University of Karlsruhe (TH)
Department of Mathematics
Institute of Stochastics

Script

Applied Stochastic Models

Contents

Nomenclature

$\mathbf{N}$	Natural numbers (including 0)
$\mathbf{Z}$	Integers
$\mathbf{Q}$	Rationals
$\mathbf{R}$	Reals
$\mathbf{C}$	Complex numbers
$\Re$	real part
$\Im$	imaginary part
P	probability measure
$P(A)$	the probability of A
$E[X]$	the expectation of X
$\mathrm{Var}(X)$	the variance of X
P^X	pushforward measure of X, see page 13
$B(m,n)$	beta distribution with mean $\frac{m}{m+n}$
$\mathrm{BIN}(n,p)$	binomial distribution with mean np and variance $np(1-p)$
$\chi^2(n)$	Chi square distribution with n degrees of freedom
$\mathrm{DU}(1,\dots,n)$	discrete uniform distribution over the set $\{1,\dots,n\}$
$\mathrm{ERL}_n(\lambda)$	Erlang distribution
$\mathrm{EXP}(\alpha)$	exponential distribution with mean $\frac{1}{\alpha}$
$\Gamma(n,\alpha)$	gamma distribution with mean $\frac{n}{\alpha}$ and variance $\frac{n}{\alpha^2}$
$\mathrm{GEOM}(p)$	geometric distribution with mean $\frac{1}{p}$
$\mathcal{N}(\mu,\sigma^2)$	normal distribution with mean μ and variance σ^2
$\mathrm{NB}(n,p)$	negative binomial distribution with mean $n\frac{1-p}{p}$ and variance $n\frac{1-p}{p^2}$
$\mathrm{POI}(\lambda)$	Poisson distribution with mean λ
$\mathrm{UNIF}(a,b)$	uniform distribution over the intervall $[a,b]$
$\mathrm{WEI}(\lambda,\alpha)$	Weibull distribution
CLT	central limit theorem
SSLN	strong law of large numbers
WLLN	weak law of large numbers
a. e.	almost everywhere
cdf	comulated density function

e. g.	exampli gratia, for example
i. e.	id est, that is
iff	if and only if
iid	independent identically distributed
pdf	probability density function

Chapter 1

Preliminaries to Probability Theory

1.1 Probability

1.1.1 Problems

1) A men with n different keys wants to open his door. Find the probability distribution of the number or attempts he will need to open the door, if unsuccessful keys are eliminated.

Let X be the number of attempts.

$$
\begin{aligned}
P(X=1) &= \frac{1}{n} \\
P(X=2) &= \frac{n-1}{n}\frac{1}{n-1} = \frac{1}{n} \\
P(X=3) &= \frac{n-1}{n}\frac{n-2}{n-1}\frac{1}{n-2} = \frac{1}{n} \\
&\vdots \\
P(X=k) &= \frac{n-1}{n}\frac{n-2}{n-1}\cdots\frac{n-k+1}{n-k+2}\frac{1}{n-k+1} = \frac{1}{n}
\end{aligned}
\tag{1.1}
$$

So X follows a discrete uniform distribution $\mathrm{DU}(1, 2, \ldots, n)$.

$$\mathrm{E}[X] = \frac{n+1}{2} \tag{1.2}$$

$$\mathrm{E}[X^2] = \frac{(n+1)(2n+1)}{6} \tag{1.3}$$

$$\mathrm{Var}(X) = \frac{n^2-1}{12} \tag{1.4}$$

2) What happens if unsuccessful keys are not eliminated?

$$\mathrm{P}(X = k) = \left(1 - \frac{1}{n}\right)^{k-1} \frac{1}{n}, \quad k = 1, 2, \ldots \tag{1.5}$$

So X follows a geometric distribution $\mathrm{GEOM}(\frac{1}{n})$.

$$\mathrm{E}[X] = \frac{1}{p} = n \tag{1.6}$$

$$\mathrm{E}[X^2] = \frac{1-p}{p^2} = n(n-1) \tag{1.7}$$

3) There are N chips numbered $1, 2, 3, \ldots, N$.

Chips are drawn one by one with replacement. Let X denote the number of trials needed to get a chip drawn earlier.

Possible values are: $2, 3, \ldots, N+1$.

$$\begin{aligned}
\mathrm{P}(X > 2) &= 1 - \frac{1}{n} \\
\mathrm{P}(X > 3) &= \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \\
&\vdots \\
\mathrm{P}(X > k) &= \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right)
\end{aligned} \tag{1.8}$$

$$\mathrm{P}(X = k) = \mathrm{P}(X > k-1) - \mathrm{P}(X > k)$$

Let X be a discrete random variable on $0, 1, 2, \ldots$ with $P(X = k) = p_k$.

$$\begin{aligned} E[X] &= \sum_{n=1}^{\infty} n p_n = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} 1 \right) p_n = \sum_{k=1}^{\infty} \left(\sum_{n=k}^{\infty} p_n \right) \\ &= \sum_{k=1}^{\infty} P(X \geqslant k) = \sum_{k=0}^{\infty} P(X > k) \end{aligned} \tag{1.9}$$

If X is a non-negative continuous random variable with cdf $F(x)$, then

$$E[X] = \int_0^{\infty} [1 - F(x)] dx \tag{1.10}$$

For $X_c = \min\{X, c\} = \begin{cases} X & X \leqslant c \\ c & X > c \end{cases}$

$$E[X_c] = \int_0^{c} [1 - F(x)] dx \tag{1.11}$$

$$\begin{aligned} E[X] = 2 + \left(1 - \frac{1}{N}\right) + \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) + \cdots \\ \cdots + \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{N-1}{N}\right) \end{aligned} \tag{1.12}$$

4) Draw k chips without replacement.

Let Y denote the maximum label on a chip in the sample. $Y = k, \ldots, N$.

$$P(Y = n) = \frac{\binom{n-1}{k-1}}{\binom{N}{k}} \tag{1.13}$$

To see this, remember the recursion

$$\binom{N}{k} = \binom{N-1}{k-1} + \binom{N-1}{k} \tag{1.14}$$

and obtain by induction

$$\binom{N}{k} = \binom{N-1}{k-1} + \binom{N-2}{k-1} + \cdots + \binom{k-1}{k-1} + \underbrace{\binom{k-1}{k}}_{0} \tag{1.15}$$

So this really is a probability distribution. Furthermore, there are $\binom{N}{k}$ ways to choose k chips form N chips. In this k chosen chips there has to be chip number n and over this $k-1$ further chips from the $n-1$ chips less then n. Therefore exists $\binom{n-1}{k-1}$ ways to realize this.

$$\mathbf{E}[Y] = \sum_{n=k}^{N} n\,\mathbf{P}(Y=n) = \sum_{n=k}^{N} n\frac{\binom{n-1}{k-1}}{\binom{N}{k}} = \sum_{n=k}^{N} k\frac{\binom{n}{k}}{\binom{N}{k}} = k\frac{\binom{N+1}{k+1}}{\binom{N}{k}} = k\frac{N+1}{k+1} \tag{1.16}$$

Here, beside the formula from above to calculate the sum, it was two times used that $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$.

On the same way calculate

$$\mathbf{E}[Y(Y+1)] = \cdots = \sum_{n=k}^{N} k(k+1)\frac{\binom{n+1}{k+1}}{\binom{N}{k}} = \cdots = k(N+1)\frac{N+2}{k+2} \tag{1.17}$$

And from this

$$\mathbf{Var}(Y) = \mathbf{E}[Y(Y+1)] - \mathbf{E}[Y]\,\mathbf{E}[Y+1] = \cdots = \frac{k(N+1)(N-k)}{(k+1)^2(k+2)} \tag{1.18}$$

5) Draw $2n+1$ chips without replacement.

Let Z denote the median of the sample.

Observe that

$$\mathbf{P}(Z=k) = \frac{\binom{k-1}{n}\binom{N-k}{n}}{\binom{N}{2n+1}} \tag{1.19}$$

Since $\mathbf{P}(Z=k) = \mathbf{P}(Z=N-k+1)$

$$\mathbf{E}[Z] = \frac{N+1}{2} \tag{1.20}$$

Analogous to the calculations from the maximum of the sample obtain that

$$\mathbf{E}[Z(N-Z+1)] = \frac{(n+1)(N+1)(N+2)}{2(2n+3)} \tag{1.21}$$

and so

$$\mathbf{Var}(Z) = \frac{(N+1)(N-2n-1)}{8n+12} \tag{1.22}$$

1.2 Background

Discrete Distributions: Bernoulli, binomial, Poisson, geometric, negative binomial, hyper geometric.

Continuous distributions: uniform, exponential, gamma, normal, beta, Weibull.

1. If $X_1, \ldots, X_n$ are independent Bernoulli variables then $X_1 + \cdots + X_n \sim \mathrm{BIN}(n, p)$.

2. If $X \sim \mathrm{BIN}(n, p)$, $n \to \infty$, $p \to 0$ so that $np \to \lambda$ $(0 < \lambda < \infty)$, then $X \to \mathrm{POI}(\lambda)$.

3. If $X_1, \ldots, X_n$ are iid $\mathrm{EXP}(\alpha)$ then $X_1 + \cdots + X_n \sim \Gamma(n, \alpha)$.

4. If $X_1, \ldots, X_n$ are iid $\mathrm{GEOM}(p)$ then $X_1 + \cdots + X_n \sim \mathrm{NB}(n, p)$.

5. If X_i's are independent $\mathcal{N}(\mu_i, \sigma_i^2)$ then $\sum_i X_i \sim \mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$.

6. If $X_i, \ldots, X_n$ are iid with finite mean μ and finite variance σ^2 then $\bar{X}$ is asymptotically $\mathcal{N}(\mu, \sigma^2)$. (CLT)

7. $X \sim \mathrm{B}(m, n)$ if it has pdf

$$f(x) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} x^{m-1}(1-x)^{n-1}, \quad 0 < x < 1 \tag{1.23}$$

8. $X \sim \mathrm{WEI}(\lambda, \alpha)$ if it has pdf

$$f(x) = \lambda\alpha t^{\alpha-1} e^{-\lambda t^{\alpha}}, \quad t > 0,\ \alpha > 0,\ \lambda > 0 \tag{1.24}$$

1.2.1 Formulas

1.

$$\frac{1}{1-z} = 1 + z + z^2 + \dots \qquad |z| < 1 \tag{1.25}$$

$$\frac{1}{(1-z)^2} = 1 + 2z + 3z^2 + \dots \qquad |z| < 1 \tag{1.26}$$

$$\frac{2}{(1-z)^3} = 1\cdot 2 + 2\cdot 3z + 3\cdot 4z^2 + \dots, \qquad |z| < 1 \tag{1.27}$$

2.

$$\int_0^\infty e^{-\alpha t} t^{n-1} dt = \frac{\Gamma(n)}{\alpha^n}, \quad \alpha > 0, n > 0 \tag{1.28}$$

$$\Gamma(n+1) = n\Gamma(n) \tag{1.29}$$

9. If X has cdf $F(x) = P(X \leqslant x)$ and X is continuous then $F(X) \sim \mathbf{UNIF}(0,1)$

   Let $Y = F(X)$ then $P(Y \leqslant y) = P(F(X) \leqslant y) = P(X \leqslant F^{-1}(y)) = F(F^{-1}(y)) = y$, $0 < y < 1 \Longrightarrow F(X) \sim \mathbf{UNIF}(0,1)$.

   If $X \sim \mathbf{EXP}(\alpha)$ then $F(x) = 1 - e^{-\alpha x}$, $x > 0$ and $F(X) = 1 - e^{-\alpha X} \sim \mathbf{UNIF}(0,1)$.

$$\begin{aligned} U &\sim \mathbf{UNIF}(0,1) \\ 1 - e^{-\alpha X} &= U \\ e^{-\alpha X} &= 1 - U \\ -\alpha X &= \ln(1-U) \\ X &= -\tfrac{1}{\alpha}\ln(1-U) \cong -\tfrac{1}{\alpha}\ln(U) \end{aligned} \tag{1.30}$$

10. If $U \sim \mathbf{UNIF}(0,1)$ then U and $1-U$ are identically distributed.

$$P(1-U \leqslant x) = P(U > 1-x) = 1 - P(U \leqslant 1-x) = 1 - (1-x) = x \tag{1.31}$$

11. If $X \sim \mathbf{UNIF}(a,b)$ then $\frac{X-a}{b-a} \sim \mathbf{UNIF}(0,1)$.

12. If X has pdf $f(x)$ then X^2 has pdf $\frac{1}{2\sqrt{x}}(f(\sqrt{x}) + f(-\sqrt{x}))$.

$$P(X^2 \leqslant x) = P(-\sqrt{x} \leqslant X \leqslant \sqrt{x}) = F(\sqrt{x}) - F(-\sqrt{x}) \tag{1.32}$$

pdf of X^2 is obtained as above.

If X is $\mathcal{N}(0,1)$ then pdf of X^2 is $\frac{1}{\sqrt{2\pi x}}e^{-\frac{x}{2}}$. Note that this is a $\Gamma(\frac{1}{2},\frac{1}{2})$. $\left(\frac{\alpha^n}{\Gamma(n)}x^{n-1}e^{-\alpha x}\right)$

If $X_1,\dots,X_n$ are iid $\mathcal{N}(0,1)$ then $X_1^2+\cdots+X_n^2 \sim \Gamma(\frac{n}{2},\frac{1}{2}) \quad (\cong \chi^2(n))$.

13. If $X_1,\dots,X_n$ are iid $\mathrm{EXP}(\alpha)$ then $X_1+\frac{X_2}{2}+\cdots+\frac{X_n}{n}$ and $\max\{X_1,\dots,X_n\}$ are identically distributed. See also [BP87] and [BP81].

Proof: Let $Y_n=\max(X_1,\dots,X_n)$, then

$$\begin{aligned} P(Y_n\leqslant y) &= P(\max(X_1,\dots,X_n)\leqslant y) = P(X_1\leqslant y,\dots,X_n\leqslant y)\\ &= Pr(X_1\leqslant y)\cdots P(X_n\leqslant y) = (1-e^{-\alpha y})^n \end{aligned} \tag{1.33}$$

[If the $X_i\sim \mathrm{EXP}(\alpha_i)$ not identically distributed but just independent exponential then $P(Y\leqslant y)=\prod_{i=1}^n(1-e^{-\alpha_i y})$ (Euler's pentagonal numbers).]

Let $Z_n=X_1+\frac{X_2}{2}+\cdots+\frac{X_n}{n}$. When $n=1$, $Z=Y=X_1$ and the claim is true. Assume the claim is true for $n=m$, then

$$P\Big(\frac{X_{m+1}}{m+1}\leqslant y\Big) = P(X_{m+1}\leqslant (m+1)y) = 1-e^{-\alpha(m+1)y} \tag{1.34}$$

and the pdf of $\frac{X_{m+1}}{m+1}$ is $\alpha(m+1)e^{-\alpha(m+1)y}$ for $y>0$.

$$\begin{aligned} P(Z_m\leqslant x) &= P\Big(\Big(X_1+\frac{X_2}{2}+\cdots+\frac{X_m}{m}\Big)+\Big(\frac{X_{m+1}}{m+1}\Big)\leqslant x\Big)\\ &= \int_0^\infty P\Big(X_1+\frac{X_2}{2}+\cdots+\frac{X_m}{m}\leqslant x-y\,\Big|\,\frac{X_{m+1}}{m+1}=y\Big)\,P^{\frac{X_{m+1}}{m+1}}(dy)\\ &= \int_0^x (1-e^{-\alpha(x-y)})^m\alpha(m+1)e^{-\alpha(m+1)y}dy\\ &= \int_0^x (m+1)(e^{-\alpha y}-e^{-\alpha x})^m\alpha e^{-\alpha y}dy\\ &= \int_{e^{-\alpha x}}^1 (m+1)(u-e^{-\alpha x})^m du = [(u-e^{-\alpha x})^{m+1}]_{u=e^{-\alpha x}}^1\\ &= (1-e^{-\alpha x})^{m+1} \end{aligned} \tag{1.35}$$

Here, the substitution $u=e^{-\alpha y}$ was made to evaluate the integral. The second line is the law of total probability in the form $P(X\leqslant x)=\int P(X\leqslant x|Y=y)\,P^Y(dy)$ where P^Y denotes the pushforward measure of Y. Loosely spoken this means that $P^Y(dy)$ is $P(y\leqslant Y<y+dy)$ ($\approx P(Y=y)$) or if Y has a density,

then $g(y)dy$ where g is the pdf of Y. In this case $\frac{X_{m+1}}{m+1}$ has a density which is given by (1.33).

Thus, by induction $P(Z_n \leqslant x) = (1 - e^{-\alpha x})^n = P(Y_n \leqslant x)$ for all $n \geqslant 1$ what shows the theorem. Note especially that the expectation of Z_n is easy to calculate and also holds for Y_n. Thus one gets as corollary

$$E[\max(X_1, \ldots, X_n)] = \tfrac{1}{\alpha}(1 + \tfrac{1}{2} + \cdots + \tfrac{1}{n}) \tag{1.36}$$

for iid $X_i \sim \mathrm{EXP}(\alpha)$.

14. $U_1, U_2, \ldots$ are iid $\mathrm{UNIF}(0,1)$ [ladder indices]

$N \sim \min\{U_n > U_{n-1}, n \geqslant 2\}$

$$\begin{aligned} N = 2 &\Longrightarrow U_2 > U_1 \\ N > 2 &\Longrightarrow U_1 > U_2 \\ N > 3 &\Longrightarrow U_1 > U_2 > U_3 \\ &\vdots \\ N > n &\Longrightarrow U_1 > U_2 > \cdots > U_n \end{aligned} \tag{1.37}$$

There are $n!$ permutations of $U_1, \ldots, U_n$, so

$$P(N > n) = \frac{1}{n!} \tag{1.38}$$

$M(x) = \min\{U_1 + \cdots + U_n > x, n \geqslant 1\}, \quad 0 \leqslant x \leqslant 1$

$$\begin{aligned} P(M(x) = 1) &= P(U_1 > x) \\ P(M(x) > 1) &= P(U_1 \leqslant x) \\ P(M(x) > 2) &= P(U_1 + U_2 \leqslant x) \\ &\vdots \\ P(M(x) > n) &= P(U_1 + \cdots + U_n \leqslant x) \\ P(M(x) > n+1) &= P(U_1 + \cdots + U_{n+1} \leqslant x) \\ &= \int_0^1 P(U_1 + \cdots + U_n + y \leqslant x)\, P^{U_{n+1}}(dy) \\ &= \int_0^x P(M(x-y) > n) dy \end{aligned}$$

Here U_{n+1} has pdf 1 on $[0,1]$ and thus $P^{U_{n+1}}$ is just the Lebesgue measure, i. e. loosely spoken $P^{U_{n+1}}(dy) = dy$.

$$\begin{aligned} &\text{But} && P(M(x) > 1) = x \\ &\implies && P(M(x) > 2) = \frac{x^2}{2} \\ & && P(M(x) > 3) = \frac{x^3}{3!} \\ & && \vdots \\ &(1.39) && P(M(x) > n) = \frac{x^n}{n!} \end{aligned}$$

If $x = 1$, $M = M(1)$ then $P(M > n) = \frac{1}{n!}$. Thus M and N are identically distributed and

$$(1.40) \qquad E[M] = E[N] = \sum_{n=1}^{\infty} P(N > n) = \sum_{n=1}^{\infty} \frac{1}{n!} = e$$

15. If X and Y are independent $\mathrm{UNIF}(0,1)$ find the pdf of $Z = X + Y$.

$$(1.41) \qquad h(z) = \int_0^z f(z-y)g(y)dy \qquad \text{(Convolution)}$$

$f(x) = 1 \quad 0 < x < 1$, $g(x) = 1 \quad 0 < x < 1$. $\therefore 0 < z - y < 1$, $0 < y < 1$. $\Rightarrow$ $0 < y < z < 1 + y$. $\Rightarrow 0 < z < 2$, $\max\{0, z-1\} < y < \min\{z, 1\}$.

$$(1.42) \qquad \begin{aligned} h(z) &= \int_0^z dy = z, && 0 < z < 1 \\ h(z) &= \int_{z-1}^1 dy = 2 - z, && 1 < z < 2 \end{aligned}$$

$$(1.43) \qquad \implies \qquad h(z) = 1 - |z-1|, \quad 0 < z < 2$$

16. X_1 and X_2 have joint pdf $f(x_1, x_2)$, $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$. It exists a unique solution $X_1 = h_1(Y_1, Y_2)$, $X_2 = h_2(Y_1, Y_2)$, if h_1 and h_2 have continuous partial derivatives with $|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} \neq 0 \ \forall x_1, x_2$.

Then the joint pdf of (Y_1, Y_2) is

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(h_1(y_1, y_2), h_2(y_1, y_2)) \cdot |J| \tag{1.44}$$

if (y_1, y_2) is in the range of (Y_1, Y_2).

Ex. 1: X_1, X_2 are iid $\mathrm{EXP}(\alpha)$, $Y_1 = X_1 + X_2$, $Y_2 = \frac{X_1}{X_2}$, $\Rightarrow X_1 = \frac{Y_1 Y_2}{1+Y_2}$, $X_2 = \frac{Y_1}{1+Y_2}$.

$$|J| = \begin{vmatrix} \frac{y_2}{1+y_2} & \frac{y_1}{(1+y_2)^2} \\ \frac{1}{1+y_2} & -\frac{y_1}{(1+y_2)^2} \end{vmatrix} = \frac{y_1}{(1+y_2)^2} \tag{1.45}$$

$$f_{Y_1,Y_2}(y_1, y_2) = \alpha^2 e^{-\alpha(x_1+x_2)}|J| = \left(\frac{\alpha}{1+y_2}\right)^2 \cdot y_1 e^{-y_1} \tag{1.46}$$

Since $f_{Y_1,Y_2}(y_1, y_2)$ can be written as a product $f_{Y_1}(y_1) f_{Y_2}(y_2)$, Y_1 and Y_2 are independent.

Ex. 2: *Box-Müller transformation*

X and Y are iid $\mathcal{N}(0,1)$, $f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$, $-\infty < x, y < \infty$.

$$\begin{aligned} R^2 &= X^2 + Y^2 & \tan\Theta &= \frac{Y}{X} \\ d &= x^2 + y^2 & \tan\vartheta &= \frac{y}{x} \\ x &= \sqrt{d}\cos\vartheta & y &= \sqrt{d}\sin\vartheta \end{aligned} \tag{1.47}$$

$$|J| = \begin{vmatrix} \frac{\cos\vartheta}{2\sqrt{d}} & -\sqrt{d}\sin\vartheta \\ \frac{\sin\vartheta}{2\sqrt{d}} & \sqrt{d}\cos\vartheta \end{vmatrix} = \frac{1}{2} \tag{1.48}$$

$$f(d,\vartheta) = \frac{1}{2\pi} \cdot \frac{1}{2} e^{-\frac{1}{2}d}, \quad d \geqslant 0, \quad 0 \leqslant \vartheta < 2\pi \tag{1.49}$$

R^2 follows an exponential distribution with parameter $\frac{1}{2}$, Θ follows a uniform distribution in $[0, 2\pi)$ and R^2 and Θ are independent.

If $(X_1, \ldots, X_n)$ are iid $\mathcal{N}(\mu, \sigma^2)$, then $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ are independent.

17. If X is a discrete random variable, say on 0,1,2,...

    $E[z^X]$ is the pgf (probability generating function) of X. [$E[\frac{1}{z^X}]$ is the z-Transform]

If X is continuous then $E[e^{itX}]$ is the characteristic function (Fourier transform). $E[e^{-sX}]$ is the Laplace transform ($X \geqslant 0$). If X has pdf $f(x)$ with $|f(x)| < ke^{\gamma t}$ for some $k, \gamma > 0$, then $\mathcal{L}_X(s) := E[e^{-sX}] = \int_0^\infty e^{-st} f(t)dt$. See also [PBM92]. Some famous correspondings:

$$
\begin{array}{ccc}
f(t) & \longleftrightarrow & F(s) \\
1 & & \frac{1}{s} \\
t^n & & \frac{n!}{s^{n+1}} \\
\cos t & & \frac{1}{1+s^2} \\
\sin t & & \frac{s}{1+s^2} \\
e^{\alpha t} & & \frac{1}{s-\alpha} \quad , \quad \Re(s) > \alpha
\end{array}
\tag{1.50}
$$

Ex.: If $X_1 \sim \mathrm{EXP}(\alpha_1)$, $X_2 \sim \mathrm{EXP}(\alpha_2)$ are independent, $\alpha_1 \neq \alpha_2$, then $E[e^{-s(X_1+X_2)}] = E[e^{-sX_1}]\,E[e^{-sX_2}] = \frac{\alpha_1}{s+\alpha_1}\frac{\alpha_2}{s+\alpha_2} = \frac{\alpha_1\alpha_2}{\alpha_2-\alpha_1}\left(\frac{1}{s+\alpha_1} - \frac{1}{s+\alpha_2}\right) \Longrightarrow$ pdf of $X_1 + X_2$ is $\frac{\alpha_1\alpha_2}{\alpha_2-\alpha_1}(e^{-\alpha_1 t} - e^{-\alpha_2 t})$. Generalize this result to $X_1 + \cdots + X_n$.

If X_1 and X_2 are identically $\mathrm{EXP}(\alpha)$, pdf of $X_1 + X_2$ is $\alpha^2 te^{-\alpha t}$, conditional pdf of X_1 given $X_1 + X_2 = s$ is

$$
\frac{\alpha e^{-\alpha u} \cdot \alpha e^{-\alpha(s-u)}}{\alpha^2 s e^{-\alpha s}} = \frac{1}{s}\,, \quad 0 < u < s
\tag{1.51}
$$

so, X_1 given $X_1 + X_2 = s$ is uniform in $(0, s)$.

18. Exponential distribution lacks memory:

$$
P(X > x + y \,|\, X > x) = P(X > y)
\tag{1.52}
$$

This is also called the Markovian property.

If $G(0) = 1$, $G(x) > 0$, $\forall x > 0$, G non increasing and $G(x+y) = G(x)G(y)$, then $G(x) = e^{-[\ln G(1)]x}$. (Functional equation)

19. If $X_1, \ldots, X_n$ are independent with cdf $F_1(x), \ldots, F_n(x)$ then

$$
P(\max\{X_1, \ldots, X_n\} \leqslant x) = \prod_{i=1}^{n} F_i(x)
\tag{1.53}
$$

$$
P(\min\{X_1, \ldots, X_n\} \leqslant x) = 1 - \prod_{i=1}^{n} (1 - F_i(x))
\tag{1.54}
$$

In particular, if $X_1, \ldots, X_n$ are $\mathrm{EXP}(\alpha_i)$, then $\min\{X_1, \ldots, X_n\}$ is $\mathrm{EXP}(\alpha_1 + \cdots + \alpha_n)$. Further, if $\alpha_1 = \cdots = \alpha_n$, then $X_1 + \frac{X_2}{2} + \cdots + \frac{X_n}{n}$ is identically distributed with $\max\{X_1, \ldots, X_n\}$.

20. There are n things of which a_1 and a_2 are identical. The number of permutations is

$$\frac{n!}{a_1!\,a_2!}\,, \quad a_1 + a_2 = n \tag{1.55}$$

If there are alike $a_1 + \cdots + a_k = n$ things, the number of permutations is

$$\frac{n!}{a_1! \cdots a_k!} \tag{1.56}$$

This fraction is also called the multinomial coefficient and this leads to the multinormal distribution (and also Dirichlet distribution).

Let $X_1, X_2, \ldots, X_n$ be iid with cdf $F(x)$ and pdf $f(x)$. Arrange them in ascending order

$$\begin{aligned} &X_{(1)} < X_{(2)} < \cdots < X_{(n)} \\ &X_{(1)} = \min\{X_1, \ldots, X_n\} \\ &X_{(n)} = \max\{X_1, \ldots, X_n\} \\ &X_{(k)} = k^{\text{th}} \text{ order statistic} \end{aligned} \tag{1.57}$$

pdf of $X_{(k)}$ is

$$\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1} f(x)[1 - F(x)]^{n-k} \tag{1.58}$$

Ex.: $X_1, \ldots, X_n$ are iid UNIF$(0,1)$, pdf of $X_{(k)}$ is

$$f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k} \tag{1.59}$$

pdf of $nX_{(k)}$ is $\frac{n(n-1)\cdots(n-k+1)}{(k-1)!} \cdot \frac{1}{n^k} \cdot \frac{x^{k-1}(1-\frac{x}{n})^n}{(1-\frac{x}{n})^k} \xrightarrow{n\to\infty} \frac{1}{(k-1)!} x^{k-1} e^{-x}$, $x > 0$. (Gamma distribution $\Gamma(k,1)$)

21. *Poisson process*

Stochastic process: collection of random variables $\{X(t) : t \geqslant 0\}$, $t \in T$ parameter set, often time. $X(t) = X(t,\omega)$ takes values in a state space.

Markov process: $P(X(t) \leqslant u | X(t_1), \ldots, X(t_n)) = P(X(t) \leqslant u | X(t_n))$ $\forall t_1 < t_2 < \cdots < t_n < t$.

If $X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$ are independent, then the process is said to have independent increments.

If $X(t+h)-X(t)$ does only depend on h and not on t or $X(t)$, then it has stationary increments.

*Poison process (*PP*):* $X(t)$, $t \geqslant 0$, $X(t) \in \mathbf{N}$ is a PP if

a) $X(0) = 0$

b) $X(t)$ has stationary independent increments

c) for h small it is

$$\begin{aligned} &P(X(t+h)-X(t)=1) = \lambda h + o(t) \\ &P(X(t+h)-X(t)=0) = 1-\lambda h + o(t) \\ &P(X(t+h)-X(t)\geqslant 2) = o(t) \end{aligned} \tag{1.60}$$

$p_n(t+h) = P(n \text{ events happened in } (0, t+h))$, for $n \geqslant 1$:

$p_n(t+h) = p_n(t)(1-\lambda h + o(h)) + p_{n-1}(t)(\lambda h + o(h)) + o(h)$

$\Longrightarrow \frac{p_n(t+h)-p_n(t)}{h} = -\lambda p_n(t) + \lambda p_{n-1}(t) + o(1)$

$\Longrightarrow \boxed{p_n'(t) + \lambda p_n(t) = \lambda p_{n-1}(t),\ n \geqslant 1}$ and for for $n = 0$:

$p_0(t+h) = p_0(t)(1-\lambda h + o(h)) \Rightarrow \boxed{p_0'(t) + \lambda p_0(t) = 0}$.

$p_0(0) = 1 \Rightarrow p_0(t) = e^{-\lambda t}$ (λ parameter), $p_n(0) = 0$, $n \geqslant 1$

$\Longrightarrow p_1(t) = \lambda t e^{-\lambda t} \Rightarrow p_2(t) = \frac{(\lambda t)^2}{2!} e^{-\lambda t} \Rightarrow \cdots \Rightarrow p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$ by induction. Note that $X(t)$ has Poisson distribution $\mathrm{POI}(\lambda t)$.

$E[X(t)] = \operatorname{Var} X(t) = \lambda t$

Let T_1 be the time of occurrence of the first event then $P(T_1 > t) = p_0(t) = e^{-\lambda t}$, therefore $P(T_1 \leqslant t) = 1 - e^{-\lambda t}$. By shifting the origin of $X(t)$ to T_1, $P(T_2 - T_1 \leqslant t) = 1 - e^{-\lambda t}$. This following one can see that in general $T_n - T_{n-1}$ is exponential distributed, i. e. inter arrival times are iid $\mathrm{EXP}(\lambda)$.

Remark:

1) λ can depend on n, i. e. if $X(t) = n$, then the next event happens in $(t, t+h)$ with probability $\lambda_n h + o(h)$.

2) λ can depend on t, $\lambda = \lambda(t)$. Then $p_n(t) = e^{-\int_0^t \lambda(u)du} \frac{\left(\int_0^t \lambda(u)du\right)^n}{n!}$.

3) λ can depend on both n, t. (Note that in this and the two cases above, X has not longer stationary increments.)

4) If $X(t)$ and X_2 are independent PP with parameters λ_1 and λ_2 respective, then $X_1(t) + X_2(t)$ is also PP with parameter $\lambda_1 + \lambda_2$.

5) If the events are classified as Type I and Type II with probability p and $1-p$, then $X_I(t)$ is PP with parameter λp.

6) For a PP, $E[z^{X(t)}] = e^{-\lambda t(1-z)}$

$$\begin{aligned} E[z^{X_I(t)}] &= \sum_{n=0}^{\infty} E[z^{X_I(t)} \,|\, X(t) = n]\, P(X(t) = n) \\ &= \sum_{n=0}^{\infty} E[z^{X_I(t)} \,|\, X(t) = n]\, e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda p t(1-z)} \qquad \text{(use Binomial distribution)} \end{aligned} \tag{1.61}$$

Ex.: Packets arrive at a satellite at 100/sec $=: \lambda$. $P(X(2) = 150) = \frac{200^{150}}{150!} e^{-200} \approx 0.035\,‰$.

7) $u < t$: $P(X(u) = k \,|\, X(t) = n) = \frac{P(X(u)=k)\, P(X(t-u)=n-k)}{P(X(t)=n)}$
$= \frac{\frac{(\lambda u)^k}{k!} e^{-\lambda u} \frac{(\lambda(t-u))^{n-k}}{(n-k)!} e^{-\lambda(t-u)}}{\frac{(\lambda t)^n}{n!} e^{-\lambda t}} = \binom{n}{k} \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k}$

$\Longrightarrow X(u)$ given $X(t)$ follows $\mathrm{BIN}(X(t), \frac{u}{t})$.

$E[X(u) \,|\, X(t)] = X(t)\frac{u}{t}$, hence $E[\frac{X(u)}{u} \,|\, X(t)] = \frac{X(t)}{t}$.

8) $u < t$: $\mathrm{Cov}(X(u), X(t)) = E[X(u)X(t)] - E[X(u)]\, E[X(t)]$. Events happening in disjoint intervals are independent. $E[X(u)X(t)] = E[X(u)(X(t) - X(u))] + E[X(u)^2] = E[X(u)]\, E[X(t) - X(u)] + E[X(u)^2] = \lambda u \lambda(t-u) + \lambda u + (\lambda u)^2 = \lambda u \lambda t + \lambda u \Longrightarrow \mathrm{Cov}(X(u), X(t)) = \lambda u$.

9) Given $X(t) = 1$, find the pdf of the time of occurrence of the event

$$\frac{\lambda e^{-\lambda u} e^{-\lambda(t-u)}}{\lambda t e^{-\lambda t}} = \frac{1}{t} \qquad \Longrightarrow \qquad \text{Uniform in } (0, t)\,! \tag{1.62}$$

22. *Renewal Process*

A renewal process is a sum of iid positive random variables X_i with cdf $F(x)$. This is a generalisation of a Poisson process in which the X_i would be exponential distributed. Let $S_n = X_1 + X_2 + \cdots + X_n$.

$$P(S_n \leqslant t) = P(N(t) \geqslant n) \tag{1.63}$$

where $\{N(t)\}$ is the renewal process and counts the number of events happened till t.

$$P(S_n \leqslant t) = F_n(t) = \overbrace{F(t) * F(t) * \cdots * F(t)}^{n\text{-times}} \tag{1.64}$$

where $*$ refers to the operation defined as follows

$$G(t) * F(t) = \int_0^t G(t-y)dF(y) \tag{1.65}$$

over a Lebesgue-Stieltjes integral. When F is differentiable than the integral is $\int_0^t G(t-y)f(y)dy$. And thus

$$F_1(t) = F(t) \tag{1.66}$$

$$F_2(t) = \int_0^t F(t-y)dF(y) \qquad \text{and inductively} \tag{1.67}$$

$$F_n(t) = F_{n-1}(t) * F(t) = \int_0^t F_{n-1}(t-y)dF(y) \tag{1.68}$$

$$E[N(t)] = \sum_{n=1}^{\infty} P(N(t) \geqslant n) = \overbrace{\boxed{\sum_{n=1}^{\infty} F_n(t) = M(t)}}^{\text{Renewal function}} \tag{1.69}$$

$$M(t) = F(t) + \sum_{n=2}^{\infty} F_n(t) = F(t) + \Big(\sum_{n=1}^{\infty} F_n(t)\Big) * F(t) \tag{1.70}$$

$$\implies \underbrace{\boxed{M(t) = F(t) + \int_0^t M(y-t)dF(y)}}_{\text{Renewal Equation—a Volterra Equation}} \tag{1.71}$$

In the case where F is differentiable, i. e. the X_i have a density, one can get limit estimates on M with help of the Laplace transform.

Let $m(t) = M'(t) = \frac{d}{dt}\sum_{n=1}^{\infty} F_n(t) = \sum_{n=1}^{\infty} F_n'(t) = \sum_{n=1}^{\infty} f_n(t)$ and observe that $f_n(t) = F_n'(t) = \frac{d}{dt}\int_0^t F_{n-1}(t-y)dF(y) = F_{n-1}(0)f(t) + \int_0^t F_{n-1}'(t-y)dF(t) = \int_0^t f_{n-1}(t-y)f(y)dy$ is the convolution of f_{n-1} and f. Here, $F_{n-1}(0) = 0$ because

the X_i were supposed to be positive. Thus,

$$\begin{aligned}\widehat{m}(s) &= \int_0^\infty m(t)e^{-st}dt = \int_0^\infty \sum_{n=1}^\infty e^{-st}dt = \sum_{n=1}^\infty \int_0^\infty f_n(t)e^{-st}dt \\ &= \sum_{n=1}^\infty \widehat{f}_n(s) = \sum_{n=1}^\infty \left(\widehat{f}(s)\right)^n = \frac{\widehat{f}(s)}{1-\widehat{f}(s)}\end{aligned} \tag{1.72}$$

where the functions with the hats are the Laplace transforms of the same functions without the hats. The Laplace transform is defined as $\widehat{f}(s) = \int_0^\infty f(t)e^{-st}dt$. Here, it was used that the Laplace transform maps the convolution to the product, thus $\widehat{f}_n(s) = \left(\widehat{f}(s)\right)^n$ and that $|f(s)| < 1$ for $\Re s > 0$ since than $|e^{-st}| < 1$ for all $t > 0$ and therefore $|\int_0^\infty f(t)e^{-st}dt| < \int_0^\infty f(t)dt = 1$.

$$\begin{aligned}\widehat{f}(s) &= \int_0^\infty e^{-st}f(t)dt = \int_0^\infty \sum_{n=1}^\infty (-1)^n \frac{s^n}{n!}t^n f(t)d \\ &= \sum_{n=1}^\infty (-1)^n \frac{s^n}{n!}\int_0^\infty t^n f(t)dt = \sum_{n=1}^\infty (-1)^n \alpha_n \frac{s^n}{n!} \\ &= 1 - \alpha_1 s + \alpha_2 \frac{s^2}{2} - \alpha_3 \frac{s^3}{3!} + \cdots = 1 - \alpha_1 s + o(s)\end{aligned} \tag{1.73}$$

where $\alpha_n = \int_0^\infty t^n f(t)dt$ are the moments of the X_i, esp. $\alpha_1 = E[X_i]$ for all i. Here, o denotes the small Landau-o. It refers to a function which goes even after dividing by s still to 0 as s goes to 0. This means more precise,

$$\frac{\widehat{f}(s) - 1 + \alpha_1 s}{s} \quad \longrightarrow \quad 0 \qquad \text{as } s \to 0 \tag{1.74}$$

esp. $\widehat{f}(s) \to 1$ as $s \to 0$, or $f(s) = 1 + o(1)$. Therefore,

$$\begin{aligned}\widehat{m}(s) &= \frac{\widehat{f}(s)}{1-\widehat{f}(s)} = \frac{1+o(1)}{1-(1-\alpha_1 s + o(s))} = \frac{1+o(1)}{\alpha_1 s + o(s)} \\ &= \frac{1+o(1)}{\alpha_1 s\,(1+o(1))} \sim \frac{1}{\alpha_1 s}\end{aligned} \tag{1.75}$$

i. e. $\alpha_1 s\widehat{m}(s) \to 1$ as $s \to 0$. And this implies

$$\boxed{m(t) \sim \frac{1}{\alpha_1}, \quad (t \to \infty)} \qquad \text{Tauberian Theorem} \tag{1.76}$$

A supporting idea to why this is true can be obtained by considering

$$\widehat{m}(s) = \int_0^\infty m(t)e^{-st}dt \sim \int_0^\infty \frac{1}{\alpha_1}e^{-st}dt = \frac{1}{\alpha_1 s} \tag{1.77}$$

If $F(t) = 1 - e^{-\lambda t}$, $f(t) = \lambda e^{-\lambda t} \implies \widehat{f}(s) = \frac{\lambda}{\lambda+s}$ and thus $\widehat{m}(s) = \frac{\lambda/(\lambda+s)}{1-\lambda/(\lambda+s)} = \lambda/s$. Therefore, $m(t) = \lambda = 1/\alpha_1$. In comparison to (1.76) here m is not just asymptotic $1/\alpha_1$ but even equal.

Generally for any F, $M(t) \sim \frac{t}{\alpha_1}$ as $t \to \infty$, even if F is not differentiable.

Chapter 2

Systems Structure

In this chapter we start the study of the reliability of complexer systems. A complex system consists of several components which are either working or not. Depending on which of the components are working either the whole system is working or not.

2.1 Structure Formula

The structure of a system can be described by a formula

Consider a system having n components. Each component is either functioning or has failed.

Let $x_i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ components is function} \\ 0 & \text{if } i^{\text{th}} \text{ component has failed} \end{cases}$

The vector $\mathbf{x} = (x_1, \ldots, x_n)$ is called a state vector.

Let $\phi(\mathbf{x}) = \begin{cases} 1 & \text{if the system is functioning when in state } \mathbf{x} \\ 0 & \text{if the system has failed when in state } \mathbf{x} \end{cases}$

$\phi =$ structure function of the system.

This definition looks somehow like it has something to do witch Boolean algebra. To use this, we need some observations which we will discuss briefly in the following.

Not. Consider a variable x which is either 0 or 1. We can think of this variable as a Boolean variable by identifying 0 with *false* and 1 with *true*. Then we can identify the Boolean operation *not* ($\neg$) with the map which maps x to $1 - x$. Thus,

we define:

$$\neg x := 1 - x \tag{2.1}$$

And. Consider two variables x and y which are either 0 or 1, each. Both variables are 1 if and only if the minimum of them are 1 and if and only if the product xy is 1. Thus, we can identify for this kind of variables *and* ($\wedge$) with *min* and with the product of the variables. Thus, we set:

$$x \wedge y := \min(x, y) = xy \tag{2.2}$$

Or. Consider two variables x and y which are either 0 or 1, each. Both variables are 0 if and only if the maximum of them are 0. To find a analogous rule using the product like for the case "and" we have to use de Morgans rule and get that x and y are 0 if and only if $\neg x$ and $\neg y$ are both 1. Thus, we can identify for this kind of variables *or* ($\vee$) with *max* and with the following formula:

$$x \vee y := \max(x, y) = \neg(\neg x \wedge \neg y) = 1 - ((1 - x)(1 - y)) \tag{2.3}$$

2.1 Definition: We can naturally raise this definition two more variables and vectors. Thus, for $\mathbf{x} = (x_1, \ldots, x_n)$ we set:

$$\neg \mathbf{x} := (\neg x_1, \ldots, \neg x_n) = (1 - x_1, \ldots, 1 - x_n) = \mathbf{1} - \mathbf{x} \tag{2.4}$$

$$\bigwedge \mathbf{x} := \bigwedge_i x_i = x_1 \wedge \cdots \wedge x_n = \prod_i x_i = \min\{x_1, \ldots, x_n\} =: \min(\mathbf{x}) \tag{2.5}$$

$$\begin{aligned} \bigvee \mathbf{x} &:= \bigvee_i x_i = x_1 \vee \cdots \vee x_n = \neg \prod_i \neg x_i = 1 - \prod_i (1 - x_i) \\ &= \max\{x_1, \ldots, x_n\} =: \max(\mathbf{x}) \end{aligned} \tag{2.6}$$

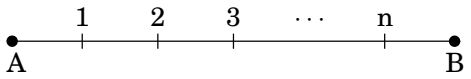
We want to reserve the symbol $\bigvee$ for the logical or or the maximum but we also want to name the construct $1 - \prod_i (1 - x_i)$. Thus, we define

$$\coprod_i x_i := 1 - \prod_i (1 - x_i) \tag{2.7}$$

and call the new symbol $\coprod$ *coproduct*.

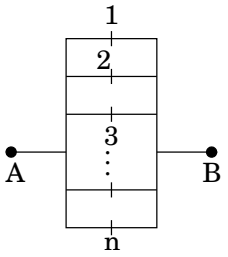
Now, let's do some examples.

Ex. 1: Series system



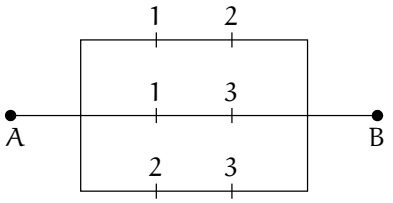
$\phi(\mathbf{x}) = x_1 \cdot x_2 \cdots x_n$. $\phi(\mathbf{x}) = 1 \Leftrightarrow x_i = 1\ \forall i$. $\phi(\mathbf{x}) = \min\{x_1, \ldots, x_n\}$.

Ex. 2: Parallel system



$\phi(\mathbf{x}) = \max\{x_1, \ldots, x_n\} = 1 - \prod_{i=1}^{n}(1 - x_i)$

Ex. 3: k-out-of-n-system

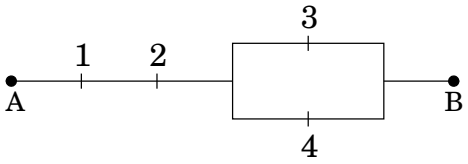


System will function if and only if at least k of the components work.

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_i x_i \geqslant k \\ 0 & \text{if } \sum_i x_i \leqslant k \end{cases} \tag{2.8}$$

Series: n-out-of-n-system
Parallel: 1-out-of-n-system

Ex. 4:



$$\phi(\mathbf{x}) = x_1 x_2 (1 - (1 - x_3)(1 - x_4)) \tag{2.9}$$

If $x_i \geqslant y_i\ \forall i$, then $\phi(\mathbf{x}) \geqslant \phi(\mathbf{y})$, i. e. ϕ is monotonically increasing.

2.1.1 Path set, cut set

For $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ we define

$$\begin{aligned} \mathbf{x} \leqslant \mathbf{y} \quad &:\Longleftrightarrow \quad x_i \leqslant y_i \quad \forall i \in \{1, \ldots, n\} \\ \mathbf{x} < \mathbf{y} \quad &:\Longleftrightarrow \quad \mathbf{x} \leqslant \mathbf{y} \quad \wedge \quad \mathbf{x} \neq \mathbf{y} \end{aligned} \tag{2.10}$$

and $>$ and $\geqslant$ in the obvious way analogously.

Let $\mathbf{x} = (x_1, \ldots, x_n)$. Partition $N = \{1, \ldots, n\}$ into A and C by $\mathbf{x}$ as follows:

$$A = \{i \in N \mid x_i = 1\} \qquad\qquad C = \{i \in N \mid x_i = 0\} \tag{2.11}$$

If $\phi(\mathbf{x}) = 1$ then $\mathbf{x}$ is a *path vector*. If further $\phi(\mathbf{y}) = 0$ for all $\mathbf{y} < \mathbf{x}$ then $\mathbf{x}$ is a *minimal path vector* and A is called a *minimal path set*. Let $A_1, \ldots, A_s$ be *all* minimal path sets to a certain system. If $\phi(\mathbf{x}) = 1$ than all components of at least one of the A_i has to work. Therefore,

$$\phi(\mathbf{x}) = \bigvee_j \bigwedge_{i \in A_j} x_i = \max_j \min_{i \in A_j} x_i = 1 - \prod_{j=1}^{s} \Big(1 - \prod_{i \in A_j} x_i\Big) = \coprod_j \prod_{i \in A_j} x_i \tag{2.12}$$

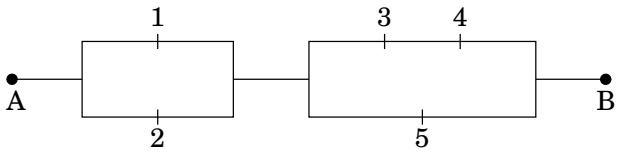
Analogously, suppose $\phi(\mathbf{x}) = 0$, then $\mathbf{x}$ is a *cut vector* and if $\phi(\mathbf{y}) = 1$ for all $\mathbf{y} > \mathbf{x}$ then $\mathbf{x}$ is a *minimal cut vector* and C is called a *minimal cut set*. Let $C_1, \ldots, C_s$ be *all* minimal cut sets to a given system. If $\phi(\mathbf{x}) = 1$ in each C_i has at least one component to work. Therefore,

$$\phi(\mathbf{x}) = \bigwedge_j \bigvee_{i \in C_j} x_i = \min_j \max_{i \in C_j} x_i = \prod_{j=1}^{s} \Big(1 - \prod_{i \in C_j} (1 - x_i)\Big) = \prod_j \coprod_{i \in C_j} x_i \tag{2.13}$$

Actually, the way to express ϕ in terms of the minimal cut sets or minimal path sets does not work for general systems but for all "normal" systems. As an example for a system where this approach doesn't work think of a system of three components which will only work if *exactly* two components are working. I. e. this is a 2-out-of-3-system which will not function if all component are functioning. A system for which this approach works is called a *coherent* system.

Ex. 5: *Minimal path set:* $\mathbf{x}$ is a path vector, if $\phi(\mathbf{x}) = 1$.

If from $\mathbf{y} < \mathbf{x}$ follows that $\phi(\mathbf{y}) = 0$, then $\mathbf{x}$ is called *minimal path vector* and $A = \{i \mid x_i = 1\}$ is called a minimal path set.



$$\begin{aligned}\phi(\mathbf{x}) &= (1-(1-x_1)(1-x_2))(1-(1-x_3x_4)(1-x_5)) \\ &= \max\{x_1,x_2\}\max\{x_3,x_4,x_5\}\end{aligned} \tag{2.14}$$

$\{1,3,4\}$, $\{1,5\}$, $\{2,3,4\}$ and $\{2,5\}$ are the minimal path sets.

Ex. 6: In k-out-of-n-system, there are $\binom{n}{k}$ minimal path sets. Let $A_1, A_2, \ldots, A_s$ denote the minimal path sets. Let

$$\begin{aligned}\alpha_j(\mathbf{x}) &= \begin{cases}1 & \text{if all the components of } A_j \text{ are functioning}\\ 0 & \text{otherwise}\end{cases} \\ &= \prod_{i\in A_j} x_i\end{aligned} \tag{2.15}$$

Then

$$\phi(\mathbf{x}) = \max_j \alpha_j(\mathbf{x}) = \max_j \prod_{i\in A_j} x_i = \coprod_j \prod_{i\in A_j} x_i \tag{2.16}$$

Any arbitrary system can be expressed as a parallel arrangement of series systems.

2.2 Reliability function

Let $f(x,\theta)$ be a density depending on an unknown parameter θ. $\hat{\theta}(x_1,\ldots,x_n)$ is called estimator for θ. If $E[\hat{\theta}] = \theta$, $\hat{\theta}$ is said to be *unbiased* and the Cramér-Rao inequality holds:

$$E[(\hat{\theta}-\theta)^2] \geqslant \frac{1}{E[(\frac{d}{d\theta}\ln f)^2]} \quad . \tag{2.17}$$

$E[(\frac{d}{d\theta}\ln f)^2]$ is also called the Fisher information or entropy of $\hat{\theta}$.

In the last section we found a function which is able to describe the structure of a system by taking values which describes which components are working and results a value which describes if the whole system is then working or not. Now,

we want to involve some probability theory. Therefore, we put in the structure function random variables which are describing whether a component is working or not. Furthermore, we assume that the random variable X_i describing whether the i^{th} component is working or not is a Bernoulli variable, i. e. X_i is 1 with a certain probability p_i and 0 otherwise and that the X_i are independent.

Consider a system which has n components. The random variables $X_1, X_2, \ldots, X_n$ describes whether the respectively components are working or not. Let $p_j := P(X_j = 1)$, $\mathbf{X} := (X_1, \ldots, X_n)$ and $\mathbf{p} = (p_1, \ldots, p_n)$. Then the *reliability* of the system is

$$r(\mathbf{p}) := P(\phi(\mathbf{X}) = 1) = E[\phi(\mathbf{X})] \tag{2.18}$$

With this we are now able to express the probability that the system is working in terms of the probabilities that the components are working. Note that the function r can only be reasonably expressed in terms of $\mathbf{p}$ when the random variables X_i are independent. Thus, in further we always think of systems with independent components, i. e. the probability that a component is functioning is independent of the state of the other components.

For further calculations we need some useful formulas. For $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ define

$$\mathbf{xy} := (x_1 y_1, \ldots, x_n y_n) \quad , \tag{2.19}$$

$$\max(\mathbf{x}, \mathbf{y}) := (\max(x_1, y_1), \ldots, \max(x_n, y_n)) \quad \text{and} \tag{2.20}$$

$$\min(\mathbf{x}, \mathbf{y}) := (\min(x_1, y_1), \ldots, \min(x_n, y_n)) \quad . \tag{2.21}$$

2.2 Theorem: *For a coherent system ϕ is increasing, i. e. $\phi(\mathbf{x}) \leqslant \phi(\mathbf{y})$ for $\mathbf{x} \leqslant \mathbf{y}$.*

Proof. A coherent system $\mathbf{x}$ is working if at least one component in each minimal path set A_i is working. If the system $\mathbf{x}$ is working and $\mathbf{x} \leqslant \mathbf{y}$ then $\mathbf{y}$ is also working since in each minimal path set is still at least one component working. Since ϕ can only take the values 0 and 1 and we have shown that if ϕ is already 1 it cannot fall back to 0 for increasing components we have already shown that ϕ is non-decreasing and therefore increasing. □

Remark: I assume that each system with increasing structure formula is also coherent but I have no proof for this.

2.3 Theorem: *If ϕ is monotone increasing ($\nearrow$) in $\mathbf{x}$ then r is $\nearrow$ in $\mathbf{p}$.*

Proof.

$$\begin{aligned} r(\mathbf{p}) &= \mathbf{E}[\phi(\mathbf{X})] = \mathbf{E}[\phi(\mathbf{X}) \mid X_i = 1]p_i + \mathbf{E}[\phi(\mathbf{X}) \mid X_i = 0](1 - p_i) \\ &= \underbrace{\mathbf{E}[\phi(\mathbf{X}) \mid X_i = 1] - \mathbf{E}[\phi(\mathbf{X}) \mid X_i = 0]}_{\geqslant 0, \text{ by monotony of } \phi} p_i + \mathbf{E}[\phi(\mathbf{X}) \mid X_i = 0] \end{aligned}$$

And therefore r is $\nearrow$ in each component p_i. □

Remark: This is especially true for coherent systems. Also for all systems we think of.

2.4 Theorem: *Redundancy on component level is better then redundancy on system level, or*

$$r(\mathbf{1} - (\mathbf{1} - \mathbf{p})(\mathbf{1} - \mathbf{p}')) \geqslant 1 - (1 - r(\mathbf{p}))(1 - r(\mathbf{p}')) \quad . \tag{2.22}$$

Here $\mathbf{1}$ denotes the *one vector* $(1, \ldots, 1)$, $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{p}' = (p'_1, \ldots, p'_n)$.

Suppose you have two identically build systems, in one the components work with probabilities $\mathbf{p}$ and in the other with $\mathbf{p}'$. On the left hand side $\mathbf{1} - (\mathbf{1} - \mathbf{p})(\mathbf{1} - \mathbf{p}')$ corresponds to a probability vector of a system in which a component works if the component works in either of both systems. On the right hand side $1 - (1 - r(\mathbf{p}))(1 - r(\mathbf{p}'))$ is the probability that either of both systems work.

Proof. On the left hand side (LHS) we have

$$\begin{aligned} r(\mathbf{1} - (\mathbf{1} - \mathbf{p})(\mathbf{1} - \mathbf{p}')) &= \mathbf{E}[\phi(\mathbf{1} - (\mathbf{1} - \mathbf{X})(\mathbf{1} - \mathbf{X}'))] \\ &= \mathbf{E}[\phi(\max(\mathbf{X}, \mathbf{X}'))] \end{aligned}$$

and on the right hand side (RHS) for independent $\mathbf{X}$ and $\mathbf{X}'$

$$\begin{aligned} 1 - (1 - r(\mathbf{p}))(1 - r(\mathbf{p}')) &= 1 - (1 - \mathbf{E}[\phi(\mathbf{X})])(1 - \mathbf{E}[\phi(\mathbf{X}')]) \\ &= \mathbf{E}[1 - (1 - \phi(\mathbf{X}))(1 - \phi(\mathbf{X}'))] \\ &= \mathbf{E}[\max(\phi(\mathbf{X}), \phi(\mathbf{X}'))] \end{aligned}$$

But $\mathbf{E}[\phi(\max(\mathbf{X}, \mathbf{X}'))] \geqslant \mathbf{E}[\max(\phi(\mathbf{X}), \phi(\mathbf{X}'))]$ since by monotony of ϕ we have that $\phi(\max(\mathbf{x}, \mathbf{x}')) \geqslant \phi(\mathbf{x})$ and $\phi(\max(\mathbf{x}, \mathbf{x}')) \geqslant \phi(\mathbf{x}')$ for all possible realisations $\mathbf{x}$ and $\mathbf{x}'$ of $\mathbf{X}$ and $\mathbf{X}'$. □

In the notation with the coproduct this can be generalized to

$$r(\coprod_i \mathbf{p}_i) \geqslant \coprod_i r(\mathbf{p}_i) \tag{2.23}$$

for $\mathbf{p}_i = (p_1^{(i)}, \ldots, p_n^{(i)})$ and $1 \leqslant i \leqslant k$. The fact that redundancy on components level is better than redundancy on system level is even true if just ϕ is a increasing function (and therefore for any coherent system if the remark to Theorem 2.2 is true), it is not necessarily needed that the components are independent. But then, of course, the result can not be expressed in terms of the reliability function.

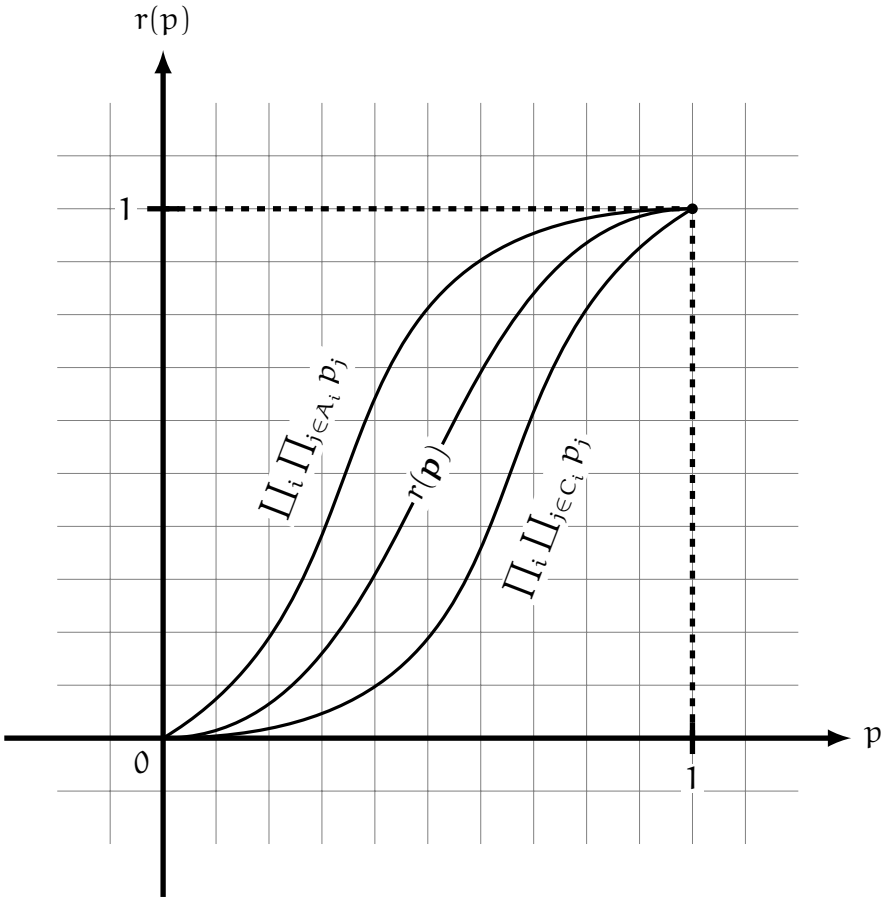
2.3 Bounds on the Reliability Function

We found easy formulas for ϕ using the minimal path and cut sets. Unfortunately, we cannot transfer this results to the reliability function.

For an example think of an 2-out-of-3-system. We can write the structure formula for this as $\phi(\mathbf{x}) = 1-(1-x_1x_2)(1-x_1x_3)(1-x_2x_3)$ but despite the components are independent e. g. the terms $1-x_1x_2$ and $1-x_1x_3$ are not since x_1 occurs in both of them.

But at least we obtain bounds for the reliability function. It holds that

$$\prod_i \coprod_{j\in C_i} p_j = \prod_i \Big(1-\prod_{j\in C_i}(1-p_j)\Big) \leqslant r(\mathbf{p}) \leqslant 1-\prod_i \Big(1-\prod_{j\in A_i} p_j\Big) = \coprod_i \prod_{j\in A_i} p_j \tag{2.24}$$


r(p)
1
0
1
p
$\prod_i \coprod_{j \in A_i} p_j$
$r(\boldsymbol{p})$
$\coprod_i \prod_{j \in C_i} p_j$

Chapter 3

System Life as a Function of Components Life

In the previous chapter we took the probabilities that the components of system are function as constant. In this chapter now we want to look at them as functions of the time. It is prudent to assume that the probability that a component is working changes over time.

Consider a system of n components. Each of the components is functioning when the system starts. Over time more and more components will fail. Let X_i be the random variable describing the life time of the i^{th} component (previously X_i was a Bernoulli variable describing whether the i^{th} component is functioning or not, this now changes). Further, let $F_i(t) = P(X_i \leqslant t)$ be the probability that the i^{th} component has already failed up to the time t and $\bar{F}_i(t) = P(X_i > t) = 1 - F_i(t)$ the probability that the i^{th} component is still functioning at time t.

We can now describe the probability $\bar{F}(t)$ that the system is functioning at time t as $\bar{F}(t) = r(\bar{F}_1(t), \ldots, \bar{F}_n(t))$ and the probability that the system has already failed at time t by $F(t) = 1 - \bar{F}(t)$. Here, r is the reliability function as in the chapter before.

$F(t)$ is the cdf of the random variable X describing the lifetime of the whole system. For the future we will assume that X is a continuous random variable and has pdf $f(t) = F'(t)$.

Then, we will define the *failure rate* $\lambda(t)$ of the system to be

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} = \frac{f(t)}{1 - F(t)} \quad . \tag{3.1}$$

Thus, the probability that the system fails in the time between t and $t + dt$ is given approximately proportional to $\lambda(t)$ for small dt, i. e. it is about $\lambda(t)dt$.

3.1 Increasing And Decreasing Failure Rates

A system has IFR (*increasing failure rate*) if $\lambda(t)$ is $\nearrow$ and it has DFR (*decreasing failure rate*) if $\lambda(t)$ is $\searrow$.

If $\lambda(t)$ is constant, the lifetime of the system is exponentially distributed.

Now, we want to look for qualitative dependencies between the failure rate of the components and the failure rate of the system. Therefore, in a first step we assume that the lifetimes of all components are identically distributed. Thus, we can write

$$\bar{F}(t) = r(\bar{F}_1(t), \ldots, \bar{F}_n(t)) = r(\bar{F}_1(t), \ldots, \bar{F}_1(t)) = r(\bar{F}_1(t)) \quad \text{and} \tag{3.2}$$

$$\begin{aligned} \lambda(t) &= \frac{\frac{d}{dt}F(t)}{\bar{F}(t)} = \frac{\frac{d}{dt}[1 - r(\bar{F}_1(t))]}{r(\bar{F}_1(t))} = \frac{r'(\bar{F}_1(t))f_1(t)}{r(\bar{F}_1(t))} \\ &= \frac{r'(\bar{F}_1(t))\bar{F}_1(t)}{r(\bar{F}_1(t))}\frac{f_1(t)}{\bar{F}_1(t)} = \left.\frac{r'(p)p}{r(p)}\right|_{p=\bar{F}_1(t)} \lambda_1(t) \quad . \end{aligned} \tag{3.3}$$

Since $\bar{F}_1(t)$ is a decreasing function of t, $\lambda(t)$ is a increasing function of t, if $\lambda_1(t)$ is a increasing function of t and $p \mapsto \frac{pr'(p)}{r(p)}$ is a decreasing function in p.

3.1.1 Hazard Function

We can also write:

$$\lambda(t) = -\frac{-f(t)}{1 - F(t)} = -\frac{\frac{d}{dt}\bar{F}(t)}{\bar{F}(t)} = -\frac{d}{dt}\ln\bar{F}(t) \tag{3.4}$$

And with

$$\Lambda(t) := \int_0^t \lambda(\tau)d\tau \tag{3.5}$$

it follows that

$$F(t) = 1 - e^{-\Lambda(t)} \quad . \tag{3.6}$$

3.1.2 Expected Lifetime

(MTBF: Mean Time Between Failure)
(MTTF: Mean Time to Total Failure)

If X is a continuous non negative random variable with cdf $F(t)$ then:

$$E[X] = \int_0^\infty (1 - F(t))dt \tag{3.7}$$

3.2 Availability

In the last section after a component has failed it was not be replaced. In this section we assume that after a component has failed after a random amount of time it will be replaced which also takes a random amount of time.

Assume that the time the i^{th} component is functioning is $EXP(\lambda_i)$ and the repair time for the i^{th} component is $EXP(\mu_i)$. $A_i(t)$ is the probability that the i^{th} component is available at time t. From the theory of stochastic processes it is known that

$$A_i(t) = \frac{\mu_i}{\lambda_i + \mu_i} + \frac{\lambda_i}{\lambda_i + \mu_i} e^{-(\lambda_i + \mu_i)t} \tag{3.8}$$

if all components are functioning at $t = 0$. Therefore, A_i goes to $\frac{\mu_i}{\lambda_i + \mu_i}$ when $t \to \infty$.

The availability of the system is then

$$A(t) = r(A_1(t), \ldots, A_n(t)) \tag{3.9}$$

and this goes to

$$A = r(A_1, \ldots, A_n) = r(\tfrac{\mu_1}{\lambda_1 + \mu_1}, \ldots, \tfrac{\mu_n}{\lambda_n + \mu_n}) \tag{3.10}$$

as t goes to infinity.

3.2.1 Statistical Interference

We consider only the exp. case. Let $X_1, \ldots, X_n$ be iid $EXP(\frac{1}{\theta})$ so that θ is the mean of each $X_i, i = 1, \ldots, n$. This random variables shell describe the lifetimes of n identical components where θ is their mean lifetime.

Make $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$. Then

$$P(nX_{(1)} > t) = e^{-\frac{t}{\theta}} \tag{3.11}$$

since $X_{(1)} = \min_{i=1}^{n} X_i \sim \text{EXP}(\frac{n}{\theta})$, i. e. $nX_{(1)} \sim \text{EXP}(\frac{1}{\theta})$. Now, shift the origin to $X_{(1)}$ then $(n-1)(X_{(2)} - X_{(1)}) \sim \text{EXP}(\frac{1}{\theta})$, and $(n-2)(X_{(3)} - X_{(2)}) \sim \text{EXP}(\frac{1}{\theta})$ analogously, a. s. o.

We want to estimate θ from the first r failures. Define

$$\begin{aligned} Y_1 &= nX_{(1)} \\ Y_2 &= (n-1)(X_{(2)} - X_{(1)}) \\ Y_3 &= (n-2)(X_{(3)} - X_{(2)}) \\ &\vdots \\ Y_r &= (n-r+1)(X_{(r)} - X_{(r-1)}) \end{aligned} \tag{3.12}$$

They are iid $\text{EXP}(\frac{1}{\theta})$.

Define τ = "total time on test" as

$$\begin{aligned} \tau &= Y_1 + Y_2 + \cdots + Y_r \\ &= nX_{(1)} + (n-1)(X_{(2)} - X_{(1)}) + \cdots + (n-r+1)(X_{(r)} - X_{(r-1)}) \\ &= \cdots = X_{(1)} + X_{(2)} + \cdots + X_{(r)} + (n-r)X_{(r)} \end{aligned} \tag{3.13}$$

This adds the lifetimes from the r broken devices together plus the time the $n-r$ still working devices were already running.

Since $\tau = \sum_{i=1}^{r} Y_i$ and $Y_i \sim \text{EXP}(\frac{1}{\theta})$, $i = 1, \ldots, n$, we have that $\tau \sim \Gamma(r, \frac{1}{\theta})$. There is a connection between the Gamma distribution and the chi-square (χ^2) distribution. If $X_1, \ldots, X_n$ are iid $\mathcal{N}(0,1)$ then $X_1 + \cdots + X_n$ follows a χ^2 distribution with n degrees of freedom. It has the Laplace Transform $(1-2s)^{-\frac{n}{2}}$ ($\chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$). Now, $\frac{2}{\theta}\tau \sim \Gamma(r, \frac{1}{\theta} \div \frac{2}{\theta}) = \Gamma(\frac{2r}{2}, \frac{1}{2}) = \chi^2(2r)$. Thus, with $\frac{2\tau}{\theta}$ one can construct confidence intervals and tests of hypothesis (for θ).

3.2.2 Maximum Likelihood Estimation (MLE)

Chapter 4

Simulation

If $X_1, X_2, \ldots$ are iid with finite mean μ then $\overline{X} \to \mu$ in prob (WLLN) and $\overline{X} \to \mu$ a. e. (SLLN).

In general it holds that $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X \mid Y]]$ and $\operatorname{Var} X = \mathrm{E}[\operatorname{Var}(X \mid Y)] + \operatorname{Var}(\mathrm{E}[X \mid Y])$.

4.1 Random Variables Generation

Let a, m be positive integers. X_0 is given, called seed.

$$X_n = aX_{n-1} \pmod{m} \tag{4.1}$$

Then $\frac{X_n}{m}$ is approximately $\mathrm{UNIF}(0,1)$ (pseudo random number).

Evaluation of Integrals

Let U be $\mathrm{UNIF}(0,1)$. $\mathrm{E}[g(U)] = \int_0^1 g(x)\,dx$.

SLLN: $\sum_{i=i}^{n} \frac{g(U_i)}{n} \to \mathrm{E}[g(U)]$ as $n \to \infty$.

Suppose we want $\int_a^b g(x)\,dx$: Put $y = \frac{x-a}{b-a}$, then $\int_a^b g(x)\,dx = (b-a)\int_0^1 g(a+(b-a)y)\,dy$. For $\int_0^\infty g(x)\,dx$ put $y = \frac{1}{x+1}$, then $\int_0^\infty g(x)\,dx = \int_0^1 \frac{g(\frac{1}{y}-1)}{y^2}\,dy$.

The same idea can be used for multiple integrals $\int_0^1\int_0^1\int_0^1 \ldots \int_0^1 g(x_1,\ldots,x_n)\,dx_1\ldots dx_n = \lim_{k\to\infty} \sum_{i=1}^{k} \frac{g(U_i^{(1)},\ldots,U_i^{(n)})}{k}$, with $U_i^{(j)} \sim \mathrm{UNIF}(0,1)$.

Estimation of π

Generate $\mathrm{UNIF}(-1,1)$ random numbers X_1, X_2. (X_1, X_2) is in the unit disk when $X_1^2 + X_2^2 \leqslant 1$. The area of the unit disk is π, the area of the square between $(-1,-1)$ and $(1,1)$ is 4. Therefore, $P(X_1^2 + X_2^2 \leqslant 1) = \frac{\pi}{4}$. Thus, generate many pairs of X_1 and X_2 and check whether $X_1^2 + X_2^2 \leqslant 1$ or not. Then π is approximately $4 \times \frac{\#\text{pairs in the unit disk}}{\#\text{pairs generated}}$.

Estimation of e

Let $U_1, U_2, \ldots$ be iid $\mathrm{UNIF}(0,1)$, $N = \min\{n : \sum_{i=1}^{n} U_i > 1\}$. Then $P(N > n) = \frac{1}{n!}$ and therefore $E[N] = \sum_{n=0}^{\infty} P(N > n) = \sum_{n=0}^{\infty} \frac{1}{n!} = e$.

4.1.1 Inverse Transform Method

(1) If X is a continuous random variable with cdf $F(x)$, then $F(X)$ is uniform distributed in $(0,1)$, i. e. $F(X) = U$ or $X = F^{-1}(U)$, where $U \sim \mathrm{UNIF}(0,1)$. Since, $P(F(X) \leqslant y) = P(X \leqslant F^{-1}(y)) = F(F^{-1}(y)) = y$, $0 < y < 1$. Therefore, pdf of $F(X)$ is 1, for $0 < y < 1$.

(2) $a + (b-a)U \sim \mathrm{UNIF}(a,b)$
$2\pi U \sim \mathrm{UNIF}(0, 2\pi)$
$2U - 1 \sim \mathrm{UNIF}(-1,1)$.

(3) If U is $\mathrm{UNIF}(0,1)$ then $U \mid U \leqslant x$ is $\mathrm{UNIF}(0,x)$.

(4) If $X \sim \mathrm{EXP}(\alpha)$, $F(x) = 1 - e^{-\alpha x} \Longrightarrow 1 - e^{\alpha X} \sim \mathrm{UNIF}(0,1) \Longrightarrow -\frac{1}{\alpha}\ln(U) \sim \mathrm{EXP}(\alpha)$, since U and $1-U$ are identically distributed.

(5) If $X \sim \mathrm{WEI}(\lambda, \alpha)$, $F(x) = 1 - e^{-(\lambda x)^{\alpha}}$, $x > 0$ and $F(X) = 1 - e^{-(\lambda X)^{\alpha}} = U \implies X = \frac{1}{\lambda}(-\ln(1-U))^{\frac{1}{\alpha}} \sim \frac{1}{\lambda}(-\ln U)^{\frac{1}{\alpha}}$.

(6) $Y \sim \mathrm{ERL}_n(\alpha)$, $Y = X_1 + \cdots + X_n$, where $X_i \sim \mathrm{EXP}(\alpha) \Longrightarrow Y = -\frac{1}{\alpha}\ln(U_1 \cdots U_n)$. For $n = 2m$ and $\alpha = \frac{1}{2}$ this leads to $\chi^2_m = -2\ln(U_1 \cdots U_{2m})$.

(7) If X and Y are independent $\mathrm{ERL}_n(1)$ and $\mathrm{ERL}_m(1)$ then $\frac{X}{X+Y}$ follows a $B(m,n)$ distribution. I. e. $\frac{-\ln(U_1 \cdots U_m)}{-\ln(U_1 \cdots U_{m+n})} \sim B(m,n)$.

(8) $F(x) = \frac{x+x^2}{2}, 0 < x < 1$, $F(X) = \frac{X+X^2}{2} = U \implies X = \frac{\sqrt{1+8U}-1}{2}$ (note sign in quadratic formula, since $0 < X < 1$).

(9) $f(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$. This is a triangular distribution.

Method 1: $Y = U_1 + U_2$, where U_1, U_2 are iid $\text{UNIF}(0,1)$.

Method 2: $F(x) = \begin{cases} \frac{x^2}{2} & 0 < x < 1 \\ 1 - \frac{(2-x)^2}{2} & 1 < x < 2 \\ 1 & x > 2 \end{cases}$

- $F(X) = \frac{X^2}{2} = U \implies X = \sqrt{2U}$, if $U < \frac{1}{2}$.
- $F(X) = 1 - \frac{(2-X)^2}{2} = U \implies X = 2 - \sqrt{2(1-U)}$, if $U > \frac{1}{2}$.

Discrete Random Variable

$P(X = x_j) = p_j = F(x_j) - F(x_j - 1) = P(F(x_j - 1) < U \leqslant F(x_j))$, $U \sim \text{UNIF}(0,1)$.

Example 1: $P(X = 0) = \frac{1}{3}$, $P(X = 1) = \frac{1}{2}$, $P(X = 2) = \frac{1}{6}$. Generate U:

- If $U < \frac{1}{3}$ then $X = 0$.
- If $\frac{1}{3} \leqslant U < \frac{5}{6}$ then $X = 1$.
- If $U \geqslant \frac{5}{6}$ then $X = 2$.

Remark: It is advantages to arrange X_j's in descending order of p_j's. ?

Example 2: $X = \lfloor kU \rfloor + 1$. $P(X = j) = P(j - 1 < kU < j) = \frac{j}{k} - \frac{j-1}{k} = \frac{1}{k}$. Thus, X is unif. in $1, 2, \ldots, k$. There are at least two methods to generate X.

Example 3: Poisson. $P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda} = \frac{\lambda}{n} P(X = n-1)$, $n = 1, 2, 3, \ldots$. Generate U:

- Is $U < e^{-\lambda} = p_0$? Yes: $X = 0$. Done.
- No: Is $U < \lambda p_0 = p_1$? Yes: $X = 1$. Done.
- No: Is $U < \frac{\lambda}{2} p_1 = p_2$? Yes: $X = 2$. Done.
- No: ...
- Is $U < \frac{\lambda}{n} p_{n-1} = p_n$? Yes: $X = n$. Done. No: ...

Box-Muller-Transform

The Box-Muller-Transform is effective way to generate normal distributed random variables from uniform ones. The idea is to look at the joint distribution of two independent standard normal variables X_1 and X_2

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{x_1^2+x_2^2}{2}} \tag{4.2}$$

and substitute this with $x_1 = \sqrt{r}\cos(\theta)$ and $x_2 = \sqrt{r}\sin(\theta)$ to

$$g(r, \theta) = \frac{1}{4\pi} e^{-\frac{r}{2}} \tag{4.3}$$

where $0 \leqslant \theta < 2\pi$ and $r \geqslant 0$. Here was used that the absolute value of determinant of the transformation matrix $|\frac{\partial(x_1,x_2)}{\partial(r,\theta)}|$ is $\frac{1}{2}$.

Now, this is the joint distribution of two independent random variables R and Θ where R is $\mathrm{EXP}(\frac{1}{2})$ and Θ is $\mathrm{UNIF}(0, 2\pi)$ and this we can already construct. Thus, we can obtain the two standard normal variables X_1 and X_2 from two uniform variables U_1 and U_2 by

$$\begin{aligned} X_1 &= \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \\ X_2 &= \sqrt{-2\ln(U_1)}\sin(2\pi U_2) \end{aligned} \tag{4.4}$$

Rawification or Polar Method There is a variant of the Box-Muller-Transformation which avoids the calculations of the sin and cos functions. Therefore, look at the two $\mathrm{UNIF}(-1, 1)$ variables $V_1 = 2U_1 - 1$ and $V_2 = 2U_2 - 1$ and calculate $S = R^2 = V_1^2 + V_2^2$. If $R \leqslant 1$ then (V_1, V_2) is as a point in $\mathbf{R}^2$ contained in the unit disc and moreover, under this condition S is $\mathrm{UNIF}(0, 1)$.

Now, set $\cos(\Theta) = \frac{V_1}{\sqrt{V_1^2+V_2^2}}$ and $\sin(\Theta) = \frac{V_2}{\sqrt{V_1^2+V_2^2}}$. We just needed the angle of (V_1, V_2) to generate the sin and cos and we need one more uniform variable to generate ln term from formula (4.4). Since the radius and the angle of (V_1, V_2) are independent we can use the random variable $S = V_1^2 + V_2^2$ for this. Thus, we obtain

$$\begin{aligned} X_1 &= \sqrt{-2\tfrac{\ln(S)}{S}}\, V_1 \qquad && \text{for } V_1^2 + V_2^2 \leqslant 1, \\ X_2 &= \sqrt{-2\tfrac{\ln(S)}{S}}\, V_2 \qquad && \text{where } S = V_1^2 + V_2^2 \\ & && \text{and } V_1, V_2 \sim \mathrm{UNIF}(-1, 1) \end{aligned} \tag{4.5}$$

When $V_1^2 + V_2^2 > 1$ they have to be rejected and generated new ones up to the condition in satisfied.

4.1.2 Acceptance-Rejection Method

Let X and Y be independent random variables with pdf $f(x)$ and $g(x)$ respectively. If there exists a constant $c \in \mathbf{R}$ such that

$$\frac{f(x)}{g(x)} \leqslant c \qquad \forall x \in \mathbf{R} \tag{4.6}$$

then the conditional pdf of Y given that $cUg(Y) \leqslant f(Y)$ for a uniform U is $f(y)$, i. e.

$$f(y) = P(Y = y \,|\, cUg(Y) \leqslant f(Y)) \tag{4.7}$$

Proof.

$$\begin{aligned} P(Y = y \,|\, cUg(Y) \leqslant f(Y)) &= \frac{P(cUg(y) \leqslant f(y))\, P(Y = y)}{P(cUg(Y) \leqslant f(Y))} \\ &= \frac{\frac{f(y)}{cg(y)} g(y)dy}{\int_{-\infty}^{\infty} \frac{f(x)}{cg(x)} g(x)dx} = \frac{f(y)dy}{\int_{-\infty}^{\infty} f(x)dx} = f(y)dy \end{aligned}$$

□

Therefore, it is possible to simulate the random variables Y and U and check if $cUg(Y) \leqslant f(Y)$. If so, set $X = Y$, otherwise take new Y and U and test again. When X is generated this way it has pdf $f(x)$, indeed.

4.2 Variance Reduction Methods

Assume we have a function g of some random variables $X_1, \ldots, X_n$ and we are interested in its expectation

$$E[g(X_1, \ldots, X_n)] \quad . \tag{4.8}$$

In most cases it is not possible to calculate this value analytically and it may be to resource expensively to approximately the real value numerically, especially when n is large.

In this cases it might be useful to simulate the expectation, i. e. the random variables X_1 up to X_n are simulated and $Y := g(X_1, \ldots, X_n)$ will be calculated from this. This will be done several times, such that lots of $Y_1, \ldots, Y_m$ are calculated. Then $\bar{Y} = \frac{Y_1 + \cdots + Y_m}{m}$ is an estimator for $E[Y]$.

The variance of $\bar{Y}$ is an indicator for the quality of the result, the lower it is the better is the estimation. It holds that $\mathrm{Var}(\bar{Y}) = \frac{\mathrm{Var}(Y)}{m}$ if the simulated Y_i are independent. Thus, principally it is possible to make the result as good as necessary by simulating enough values.

In this section we want to discuss further possibilities to reduce the variance of the estimation.

4.2.1 Antithetic Variable Technique

For two not necessarily independent sets of random variables $X_1^{(1)}, \ldots, X_n^{(1)}$ and $X_1^{(2)}, \ldots, X_n^{(2)}$ take the values $Y_1 = g(X_1^{(1)}, \ldots, X_n^{(1)})$ and $Y_2 = g(X_1^{(2)}, \ldots, X_n^{(2)})$. The variance of $\bar{Y}$ is then

$$\begin{aligned} \mathrm{Var}(\bar{Y}) &= \mathrm{Var}\left(\frac{Y_1 + Y_2}{2}\right) = \tfrac{1}{4}[\mathrm{Var}(Y_1) + 2\,\mathrm{Cov}(Y_1, Y_2) + \mathrm{Var}(Y_2)] \\ &= \frac{\mathrm{Var}(Y)}{2} + \frac{\mathrm{Cov}(Y_1, Y_2)}{2} \end{aligned} \tag{4.9}$$

and this can be reduced by choosing the $X_i^{(1)}$ and $X_i^{(2)}$ in such a way that Y_1 and Y_2 are negatively correlated.

If $Y = g(X_1, \ldots, X_n)$ is depending on random variables X_i of any kind we have discussed possibilities how to express them in terms of uniform variables U_j. Thus, we can write $Y = g(X_1, \ldots, X_n) = k(U_i, \ldots, U_m)$ for some k. We shortly write $k(\mathbf{U})$ for $k(U_1, \ldots, U_m)$.

4.1 Theorem: *If* k *is a monotone function then*

$$\mathrm{Cov}(k(\mathbf{U}), k(\mathbf{1} - \mathbf{U})) \leqslant 0 \quad .$$

Thus, if we want to simulate $Y = k(\mathbf{U})$ we can use the ideas from formula (4.9) to reduce the variance of $\bar{Y}$ when we also use $Y = k(\mathbf{1} - \mathbf{U})$ as a second realisation to a given realisation $\mathbf{U}$.

I. e. for a function depending on one variable $Y = k(U)$, instead of generating $2n$ random numbers $U_1, \ldots, U_{2n}$ we can generate n random variables $U_1, \ldots, U_n$ and use $1 - U_1, \ldots, 1 - U_n$ for the others. This results in *variance reduction* since $\mathrm{Var}(\bar{Y}) = \frac{1}{2n}\big(\mathrm{Var}\, Y + \mathrm{Cov}(k(U), k(1 - U))\big) \leqslant \frac{1}{2n}\mathrm{Var}(Y)$ by Theorem 4.1.

This approach works only for monotone k.

4.2.2 Control Variates Technique

Assume we are interested in $E[g(\mathbf{X})]$ and we already now $E[f(\mathbf{X})] = \mu$, where $\mathbf{X} = (X_1, \ldots, X_n)$. Then, for a constant $a \in \mathbf{R}$ we can look at

$$W = g(\mathbf{X}) + a(f(\mathbf{X}) - \mu) \tag{4.10}$$

and use $E[W]$ as an estimator for $E[g(\mathbf{X})]$. Then the variance of W is given by

$$\mathrm{Var}(W) = \mathrm{Var}(g(\mathbf{X})) + a^2 \mathrm{Var}(f(\mathbf{X})) + 2a \mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X})) \quad . \tag{4.11}$$

Minimizing this term by calculating $\frac{\partial \mathrm{Var}(W)}{\partial a} = 0$ leads to $a\,\mathrm{Var}(f(\mathbf{X})) + \mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X})) = 0$ or $a = -\frac{\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X}))}{\mathrm{Var}(f(\mathbf{X}))}$ and thus, we get as minimal variance for W

$$\mathrm{Var}(W) = \mathrm{Var}(g(\mathbf{X})) - \frac{\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X}))^2}{\mathrm{Var}(f(\mathbf{X}))} \leqslant \mathrm{Var}(g(\mathbf{X})) \qquad \text{for} \tag{4.12}$$

$$W = g(\mathbf{X}) - \frac{\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X}))}{\mathrm{Var}(f(\mathbf{X}))}(f(\mathbf{X}) - \mu) \tag{4.13}$$

To use this approach f has to be chosen in a way such that $E[f(\mathbf{X})]$, $\mathrm{Var}(f(\mathbf{X}))$ and $\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X}))$ are known. This is often possible but not in general.

Chapter 5

Discrete Event Simulation

5.1 Simulation of a Poisson Process

We have studied this methods to generate a Poisson distribution:

(1) Generate U:

- if $U \leqslant e^{-\lambda}$, $X = 0$
- if $e^{-\lambda} < U \leqslant \lambda e^{-\lambda}$, $X = 1$
- $\vdots$

(2) $p_{j+1} = \frac{1}{j+1} p_j$

(3) $P(X = j) = P(X_1 + \cdots + X_j \leqslant 1 < X_1 + \cdots + X_{j+1})$, where X_j's are iid $EXP(\lambda)$ ($X_j = -\frac{1}{\lambda}\ln(U)$ makes $P(X = j) = P(U_1 \cdots U_{j+1} < e^{-\lambda} \leqslant U_1 \cdots U_j)$)

The pdf of a Poisson Process $N(t)$ is $P(N(t) = j) = e^{-\lambda t}\frac{(\lambda t)^j}{j!}$, $j = 0, 1, \ldots$ If X and Y are iid $EXP(\alpha)$, $X + Y \sim \Gamma(2, \alpha)$ and

$$P(X = x \,|\, X + Y = t) = \frac{P(X = x)\,P(Y = t - x)}{P(X + Y = t)} = \frac{\alpha e^{-\alpha t}\alpha e^{-\alpha(t-x)}}{\alpha^2 t e^{-\alpha t}} = \frac{1}{t} \tag{5.1}$$

To generate the Poisson Process assume $\alpha = 1$ and proceed:

Step 1: Generate U_1, U_2.

Step 2: Take $t = -\ln(U_1 U_2)$.

Step 3: Generate U_3.

Step 4: $x = tU_3$, $y = t - x$.

This can be generalized to k variables. Generate k uniform variables. Set $t = -\ln(U_1 \cdots U_k)$. Generate another $k-1$ variables.

Given $N(t) = k$, the time of occurrence of Poisson events follows an order statistic from an uniform distribution.

Arrange $U_{(1)} \leqslant \cdots \leqslant U_{(k-1)}$ then $t(U_{(i)} - U_{(i-1)})$ are EXP variables with $U_{(k)} = t$.

Let us consider a Poisson process $N(t)$

$$P(N(t) = j) = e^{-\lambda t} \frac{(\lambda t)^j}{j!} \tag{5.2}$$

Let $T_1, T_2, \ldots$ be iid EXP(λ).

To simulate the first n events, using $T_i = -\frac{1}{\lambda} \ln U_i$, we can generate $T_1 + \cdots + T_n = -\frac{1}{\lambda} \ln(U_1 \cdots U_n)$. Then the first event occurs at $-\frac{1}{\lambda} \ln U_1$, the second at $-\frac{1}{\lambda} \ln U_1 U_2$ and so on.

5.1.1 Generation of a non-homogeneous Poisson Process

$\lambda(t)$ = arrival rate depends on t. Let arrival rate of vehicles be λ. Vehicles are of two kinds, cars and trucks, with probability p and q. Arrival rate of cars is λp. Arrival rate of trucks is λq. Suppose p depends on t, $p(t)$. $\lambda(t) = \lambda p(t)$ arrival rate of cars at time t.

Generate the process Find λ such that $\lambda(t) \leqslant \lambda$ for all t. Generate a Poisson process with rate λ as we have done earlier. Count an event which happen at time t with probability $\frac{\lambda(t)}{\lambda} \leqslant 1$. The resulting process will be a non-homogeneous Poisson process with variable rate $\lambda(t)$. If $\lambda(t)$ is close to λ then the probability of accepting an event is high.

5.2 Statistic Validation

When we generate uniform random numbers, are they "truly" random?

5.2.1 χ^2-test of goodness of fit

$$\chi^2 = \sum_i \frac{(o_i - e_i)^2}{e_i} \tag{5.3}$$

where o_i = observed frequency and e_i = expected frequency. If $\chi^2 < \chi^2_{\alpha,n-1}$, accept that data fits the distribution.

Take α. We have data of 100 random numbers $0.34, 0.90, 0.25, \ldots$ Divide $(0,1)$ as $(0,1), (0.1,0.2), \ldots, (0.9,1.0)$.

Interval	o_i	e_i	$\frac{(o_i-e_i)^2}{e_i}$
$0-0.1$	8	10	0.4
$0.1-0.2$	8	10	0.4
$0.2-0.3$	10	10	0
$0.3-0.4$	9	10	0.1
$0.4-0.5$	12	10	0.4
$0.5-0.6$	8	10	0.4
$0.6-0.7$	10	10	0
$0.7-0.8$	14	10	1.6
$0.8-0.9$	10	10	0
$0.9-1$	11	10	0.1
	100	100	3.4

Since $3.4 < 16.9 = \chi^2_{0.05,9}$ (from table), accept H_0: The data is Uniform.

5.2.2 Kolmogorov-Smirnov test

Empirical distribution Glivenko's theorem.

Take a random sample and arrange it in ascending order. Attach a probability of $\frac{1}{n}$ with each item. The corresponding cdf is called empirical distribution.

5.3 Discrete Event Simulation

For example this can be used to calculate GI/G/1-queues, i. e. queues with general iid inter arrival times and one server with general distributed service time.

Let T_i be the arrival time of the i^{th} customer where we set $T_1 = 0$, t_i the inter arrival time between the i^{th} and the $i+1^{st}$ customer and s_i the service time of the i^{th} customer. Recognize that $T_{n+1} = T_n + t_n$. The waiting time q_1 of the first customer is 0. The waiting time of the second customer can be calculated as follows: The arrival time of the second customer is T_2, but the server is busy up to the service time s_1 of the first customer. Therefore, the second customer

hasn't to wait if $s_1 \leqslant T_2$ and has to wait $q_2 = s_1 - T_2$ otherwise. With the notation $x^+ := \max(0, x)$ we write for this $q_2 = (s_1 - T_2)^+$.

In general, the waiting time q_{n+1} of the $n+1^{\text{st}}$ customer can be expressed as the difference of the arrival time plus waiting time plus service time of the n^{th} customer and the arrival time of the $n+1^{\text{st}}$ customer—or 0 if the $n+1^{\text{st}}$ customer arrives after the n^{th} customer is out of the system. This can be written as

$$q_{n+1} = (T_n + q_n + s_n - T_{n+1})^+ = (q_n + s_n - t_n)^+ \tag{5.4}$$

The distributions of the s_n and t_n are given and the q_n can be calculated recursively.

Let's make an example. Assume the inter arrival times are uniform in $[2, 3.2]$ and the service times are uniform in $[1.8, 3]$. We want to create a table looking as follows:

Cust. No.	Arrival Time	Waiting Time	Service Time	Inter Arrival Time
1	0	0	s_1	t_1
2	t_1	?	s_2	t_2
3	$t_1 + t_2$	⋮		
⋮	⋮			

We can use a small Matlab® script to do this:

```
N = 10;
Table = zeros(N,5);

n = 1;
T = 0;
q = 0;
s = 1.2 * rand + 1.8;
t = 1.2 * rand + 2;
Table(1,:) = [n T q s t];

for n = 2:N
    T = T + t;
    q = max(0,q+s-t);
    s = 1.2 * rand + 1.8;
    t = 1.2 * rand + 2;
    Table(n,:) = [n T q s t];
end
```

`N` are the number of lines in the table, the result is stored in `Table`, and `rand` gives back a uniform random variable with values in $[0, 1]$.

The result could look like this:

Cust. No.	Arrival Time	Waiting Time	Service Time	Inter Arrival Time
1	0	0	2.04	2.59
2	2.59	0	2.82	3.12
3	5.71	0	2.61	2.91
4	8.62	0	2.7	2.47
5	11.09	0.23	2.59	2.22
6	13.3	0.6	2.65	2.04
7	15.34	1.21	2.13	2.06
8	17.4	1.28	1.92	2.99
9	20.39	0.21	2.64	2.38
10	22.77	0.47	2.94	2.04

It is now possible to calculate the mean waiting time W_q. In this case this is $W_q = 0.4$. This is only an estimator for the real W_q, it is as better as more values are calculated. There are also ways to improve the result by using the techniques discussed in the last chapter like antithetic variable and control variates techniques.

Other characteristic values for the queue like the mean time in system W_s, the mean length of queue L_q, and the mean number of peoples in the system L_s can be calculated form W_q with the formulas given in the last lecture Stochastic Methods in Industry I (SMI-I): $W_s = W_q + \frac{1}{\mu}$, $L_q = \lambda W_q$, and $L_s = \lambda W_s$, where $\frac{1}{\mu}$ is the mean service time and $\frac{1}{\lambda}$ is the mean inter arrival time.

Chapter 6

Monte Carlo Simulation

6.1 Monte Carlo Markov Chain (MCMC)

$\{X_n\}$ is a Markov chain if $P(X_{n+1} = j \mid X_n = i, X_{n-1}, \ldots, X_0) = P(X_{n+1} = j \mid X_n = i) = p_{ij}$.

$P = (p_{ij})$ is transition probability matrix.

6.1.1 Chapman-Kolmogorov-Equation

$P^n = (p_{ij}^{(n)})$

- A Markov chain is said to be *irreducible* if for all i and j there exists a n so that $p_{ij}^{(n)} > 0$. I. e. each state can in principle be reached from each state. This definition applies at least for finite Markov chains.
- A Markov chain is *aperiodic* if it exists a j such that $P(X_n = j \mid X_n = j) > 0$ and $P(X_{n+1} = j \mid X_0 = j) > 0$ for some n.
- A irreducible, aperiodic Markov chain is *ergodic*.

  Then $P(X_n = j) \to \pi_j$ as $n \to \infty$ for $j = 1, \ldots, m$; $\pi P = \pi$ where $\pi = (pi_1, \ldots, \pi_n)$, and $\sum_i \pi_i = 1$.
- A Markov chain is said to be reversible if $\pi_i p_{ij} = \pi_j p_{ji}$.

We generate a reversible Markov chain which has a given stationary distribution.

There are two algorithms: Hastings Metropolis and Gibb's sampler.

Consider a set of positive numbers $b_j > 0$. Assume $B = \sum_i b_i$. Then $\{\frac{b_i}{B}\}$ defines a probability distribution $P(X = j) = \frac{b_j}{B}, j = 1, \ldots, m$. When m is large we want to simulate a reversible Markov chain with $\pi_j = \frac{b_j}{B}$

Hastings-Metropolis-Algorithm

To generate a reversible Markov Chain with given $\{\pi_1, \ldots, \pi_m\}$ stationary distribution.

$$1 \to 4 \to 2 \to 8 \to 1 \to 9 \to 15 \to \cdots \qquad \text{(this means to generate)}$$

Consider an arbitrary Markov chain with transition probability transition matrix $Q = (q_{ij})$. We want to construct a reversible Markov Chain $\{X_n\}$ with the given stationary distribution from it. If $X_n = i$ (i. e. at the n^{th} stage the chain is in state i) take the i^{th} row of Q $(q_{i1}, \ldots, q_{im})$. Take this as probability distribution and generate Y according to it.

Suppose $Y = j$. Then set $X_{n+1} = j$ or $X_{n+1} = i$ with probability α_{ij} and $1 - \alpha_{ij}$ respectively, where the α_{ij} are subject to be determined. (Thus, the M. C. changes with probability α_{ij} or remains otherwise in its old state. α_{ii} and the case where $j = i$ doesn't play a role for this considerations.)

$\{X_n\}$ has then transition probability matrix $P = (p_{ij})$, where

$$\begin{aligned} p_{ij} &= \alpha_{ij} q_{ij}, \text{ for } i \neq j \\ p_{ii} &= 1 - \sum_{i \neq j} p_{ij} \end{aligned} \tag{6.1}$$

Assume this chain is reversible and has stationary distribution $\{\pi_j\}$, i. e. $\pi_i p_{ij} = \pi_j p_{ji}$. Use (6.1) to get

$$\pi_i \alpha_{ij} q_{ij} = \pi_j \alpha_{ji} q_{ji} \tag{6.2}$$

Claim: $\alpha_{ij} = \min\{\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1\}$

- If $\alpha_{ij} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}}$ then $\alpha_{ji} = 1$.
- If $\alpha_{ij} = 1$ then $\alpha_{ji} = \frac{\pi_i q_{ij}}{\pi_j q_{ji}}$.

by (6.2). Works!

Gibb's Sampler

The Gibb's sampler is a special case of the Hastings-Metropolis-Algorithm.

Consider a multivariate random variable $\mathbf{X} = (X_1, \ldots, X_n)$. Choose i randomly of $1, \ldots, n$ equal probability $\frac{1}{n}$. Generate $X = x$ with the conditional probability distribution

$$P(X = x) = P(X_i = x \,|\, X_k = x_k, \forall k \neq i) \tag{6.3}$$

Replace $\mathbf{X}$ by $\tilde{\mathbf{X}} = (X_1, \ldots, X_{i-1}, X, X_{i+1}, \ldots, X_n)$

$$(X_1, \ldots, X_i, \ldots, X_n) \longrightarrow (X_1, \ldots, X, \ldots, X_n)$$

and repeat from this point.

This is the Hastings-Metropolis-Algorithm for a more general M. C. which can take values in an multidimensional continuous space (the above version of the HMA was phrased for a M. C. with one-dimensional discrete –even finite– state space). Furthermore, the stationary distribution is not a-priori given, thus one sets all α_{ij} to 1 and takes the stationary distribution which will be generated by $\mathbf{X}$.

Chapter 7

Statistical Quality Control

7.1 Operating Characteristic

The *operating characteristic* of an statistic test is the probability that the test is accepted given a certain true reality. To make this more precisely, recall what a statistic test is.

Assume a random variable X accords to a cdf F_θ which depends on some parameter θ. We want to test a hypothesis H_0 that $\theta \in \Theta$ for some specific range Θ versus the alternative hypothesis H_A that $\theta \notin \Theta$. E. g. let X be $\mathcal{N}(\mu, 1)$ distributed and test if $\mu \geqslant 0$. Then θ is the parameter μ and Θ is the interval $[0, \infty)$.

To make a decision rule whether to accept the hypothesis H_0 or not we introduce some test value t which is somehow depending on realizations of X and therefore depending on θ. If this test value is in some given range $t \in T$ we accept H_0, otherwise we reject H_0. In our example, t could be $\sum_{i=1}^{10} x_i$ where the x_i are ten realizations of X and T could be the interval $[-5, \infty)$, i. e. H_0 will be accepted if the sum of ten realizations of X is greater or equal to -5 and rejected otherwise.

The operation characteristic $OC(\theta)$ is now the probability $P(t \in T \mid \theta)$ that H_0 will be accepted given θ, i. e. that is $OC(\mu) = P(t \geqslant -5 \mid \mu)$ in our example. Here, we have that t is $\mathcal{N}(\mu, 10)$ distributed and therefore, one can calculate that $OC(\mu) = \Phi(\frac{5+\mu}{\sqrt{10}})$ where Φ denotes the cdf of a $\mathcal{N}(0, 1)$ distributed random variable.

7.1.1 Mistakes on Statistical Tests

When one does a statistical test there will occur exactly one of four different results. Either, the hypothesis one wants to test is right and the test tells that it is right, too. Or, the hypothesis is wrong and the test says it is wrong, too. This are the two "good" cases, where the test was right and therefore all is fine. But there are also two "bad" cases, where the test is wrong. On the one hand, the test can

say that something is right, but it is not. And on the other hand, it can say that something is wrong, but it is right. The last mistake is called the α-error and is usual more critical than the other one, called the β-error.

The four cases can be collected in a tabular as follows:

	$\theta \in \Theta$	$\theta \notin \Theta$
$t \in T$	✓	β
$t \notin T$	α	✓

The most possible α-error is $\sup_{\theta\in\Theta} P(t \notin T \mid \theta) = 1 - \inf_{\theta\in\Theta} OC(\theta)$ and the most possible β-error is analogous $\sup_{\theta\notin\Theta} OC(\theta)$. In the example we have –since Φ is monotone– that the most possible α-error is $1 - \Phi(\sqrt{2.5}) \approx 0.057$. To be precise, the α- and β-error depend on the parameter θ, but it is usual to write also α for the maximal α-error and β for the maximal β-error. Thus, in our example one would write $\alpha \approx 0.057$ since it is the most possible α-error, and since $OC(\mu)$ is monotone and continuous we further have that $\beta = 1 - \alpha \approx 0.943$ is the most possible β-error.

The best possible OC-function would be the characteristic function on Θ (χ_Θ) which means that H_0 is accepted if $\theta \in \Theta$ (if H_0 is true) and rejected otherwise (if H_0 is false, or H_A is true). But this is usually impossible to reach. But the closer OC is to χ_Θ the better the test is. Thus, one can use OC to make statements about the quality of a statistical test.

7.2 Statistic Quality Control (S.Q.C.)

- Six-sigma (6σ) limit
- Evolutionary Operations (EVOP)
- Control Charts
- Acceptance Quality

A product is never good enough and should be continually improved. Improve the product in design, production and assembly stages rather that after it is produced.

If X is $\mathcal{N}(\mu, \sigma^2)$, $P(-1.96 < \frac{X-\mu}{\sigma} < 1.96) = 0.95$. I. e. $X \approx \mu \pm 2\sigma$, thus about 95% of the items fall within the two sigma limit. Analogously, all but three items in a million fall within the six sigma limit.

Ex: Suppose X is $\mathrm{EXP}(\alpha)$. Assume $x > c$, then

$$P(X > x \mid X > c) = \frac{P(X > x, X > c)}{P(X > c)} = \frac{P(X > x)}{P(X > c)} = \frac{e^{-\alpha x}}{e^{-\alpha c}} \tag{7.1}$$
$$= e^{-\alpha(x-c)} = P(X > x - c) = P(X + c > x)$$

Thus, $X \mid X > c$ and $X + c$ are identically distributed.

Consider $\mathbf{X} = (X_1, \ldots, X_n)$, X_i are independent exponential distributed with parameter λ_i. $S = \sum_{i=1}^{n} X_i$. Generate $\mathbf{X}$ such that $\sum X_i > c$.

$$f(x_1, \ldots, x_n) = \frac{\prod(\lambda_i e^{\lambda_i x_i})}{P(S > c)} \quad , \qquad \text{if } \sum X_i > c \tag{7.2}$$

Choose randomly from $1, \ldots, n$ using $\lfloor nU \rfloor + 1$, say i. Generate an $\mathrm{EXP}(\lambda_i)$ variable X s. t. $X + \sum_{j \neq i} X_j > c$.

$$(X_1, \ldots, X_i, \ldots, X_n) \longrightarrow (X_1, \ldots, X, \ldots, X_n)$$

using (7.1): $X = Y + (c - \sum_{j \neq i} X_j)^+$, where $Y \sim \mathrm{EXP}(\lambda_i)$. ... Repeat ...

Bibliography

[BP81] RE Barlow and F Proschan. *Statistical Theory of Reliability and Life Testing: Probability Models*. University of Michigan, 1981.

[BP87] RE Barlow and F Proschan. *Mathematical Theory of Reliability*. SIMA, 1987.

[Joh94] RA Johnson. *Miller and Freund's Probability and Statistics for Engineers, Fifth Edition*. Prentice-Hall, 1994.

[PBM92] AP Prudnikov, YA Brychkov, and OI Marichev. *Integrals and Series, Volume 4: Direct Laplace Transforms*. Gordon and Breach Science Publishers, 1992.

[Ros00] SM Ross. *Introduction to Probability Models, Seventh Edition*. Academic Press, 2000.

[Ros02] SM Ross. *Simulation, Third Edition*. Academic Press, 2002.

[SYF59] M Sasieni, A Yaspan, and L Friedman. *Operations Reasearch: Methods and Problems*. John Wiley & Sons, 1959.