

Formulas for probability theory and linear models SF2941

These pages (+ Appendix 2 of Gut) are permitted as assistance at the exam.

11 maj 2008

- Selected formulae of probability
- Bivariate probability
- Transforms
- Multivariate normal distribution
- Convergence
- Poisson process
- Series Expansions and Integrals

1 Probability

1.1 Two inequalities

- $A \subseteq B \Rightarrow P(A) \leq P(B)$.
- $P(A \cup B) \leq P(A) + P(B)$

1.2 Change of variable in a probability density

If \mathbf{X} has the probability density $f_{\mathbf{X}}(\mathbf{x})$, $Y = A\mathbf{X} + \mathbf{b}$, and A is invertible, then \mathbf{Y} has the probability density

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det A|} f_{\mathbf{X}}(A^{-1}(\mathbf{y} - \mathbf{b}))$$

2 Continuous bivariate distributions

2.1 Bivariate densities

2.1.1 Definitions

The bivariate vector $(X, Y)^T$ has a continuous joint distribution with density $f_{X,Y}(x, y)$ if

$$P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) du dv.$$

where

- $f_{X,Y}(x, y) \geq 0$,
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1$

Marginal distribution:

- $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy$,
- $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx$.

Distribution function

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du.$$

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

Conditional densities:

- $X | Y = y$,

$$f_{X|Y=y}(x) := \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

if $f_Y(y) > 0$.

- $Y | X = x$

$$f_{Y|X=x}(y) := \frac{f_{X,Y}(x, y)}{f_X(x)},$$

if $f_X(x) > 0$.

Bayes' formula

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{f_{Y|X=x}(y) \cdot f_X(x)}{f_Y(y)} = \\ &= \frac{f_{Y|X=x}(y) \cdot f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X=x}(y) f_X(x) dx}. \end{aligned}$$

$f_{X|Y=y}(x)$ is a *a posteriori* density for X and $f_X(x)$ is a *priori density* for X .

2.1.2 Independence

X and Y are *independent* iff

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \text{ for every } (x, y).$$

2.1.3 Conditional density of X given an event B

$$f_{X|B}(x) = \begin{cases} \frac{f_X(x)}{P(B)} & x \in B \\ 0 & \text{elsewhere} \end{cases}$$

2.1.4 Normal distribution

If X has the density $f(x; \mu, \sigma)$ defined by

$$f_X(x; \mu, \sigma) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

then $X \in N(\mu, \sigma^2)$. $X \in N(0, 1)$ is standard normal with density

$$\phi(x) = f_X(x; 0, 1).$$

The cumulative distribution function of $X \in N(0, 1)$ is for $x > 0$

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{2} + \int_0^x \phi(t) dt$$

and

$$\Phi(-x) = 1 - \Phi(x).$$

2.1.5 Numerical computation of $\Phi(x)$

Approximative values of the cumulative distribution function of $X \in N(0, 1)$, $\Phi(x)$, can be calculated for $x > 0$ by

$$\Phi(x) = 1 - Q(x), \quad Q(x) = \int_x^\infty \phi(t)dt,$$

where we use the following approximation¹:

$$Q(x) \approx \left(\frac{1}{\left(1 - \frac{1}{\pi}\right)x + \frac{1}{\pi}\sqrt{x^2 + 2\pi}} \right) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

2.2 Mean and variance

The expectations or means $E(X), E(Y)$ are defined (if they exist) by

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx,$$
$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy,$$

respectively. Variances $\text{Var}(X), \text{Var}(Y)$ are defined as

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) dx,$$
$$\text{Var}(Y) = \int_{-\infty}^{+\infty} (y - E(Y))^2 f_Y(y) dy,$$

respectively. We have

$$\text{Var}(X) = E(X^2) - (E(X))^2.$$

The *function of a random variable* $g(X)$ has expectation

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx.$$

¹P.O. Börjesson and C.E.W. Sundberg: Simple Approximations of the Error Function $Q(x)$ for Communication Applications. *IEEE Transactions on Communications*, March 1979, pp. 639–643.

2.3 Chebyshev's inequality

$$P(|X - E(X)| > \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

2.4 Conditional expectations

The conditional expectations of X given $Y = y$ is

$$E(X|Y = y) := \int_{-\infty}^{+\infty} x f_{X|Y=y}(x) dx.$$

This can be seen as $y \mapsto E(X|Y = y)$, as a function of Y .

$$\begin{aligned} E(X) &= E(E(X|Y)), \\ \text{Var}(X) &= \text{Var}(E(X|Y)) + E(\text{Var}(Y|X)). \\ E[(Y - g(X))^2] &= E[\text{Var}[Y|X]] + E[(E[Y|X] - g(X))^2]. \end{aligned}$$

2.5 Covariance

$$\begin{aligned} \text{Cov}(X, Y) &:= E(XY) - E(X) \cdot E(Y) = \\ &= E([X - E(X)][Y - E(Y)]) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - E(X))(y - E(Y)) f_{X,Y}(x, y) dx dy. \end{aligned}$$

We have

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n-1} a_i a_j \text{Cov}(X_i, X_j), \\ \text{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, X_j). \end{aligned}$$

2.6 Coefficient of correlation

Coefficient of correlation between X and Y is defined as

$$\rho := \rho_{X,Y} := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}.$$

3 Best linear prediction

α and β that minimize

$$E [Y - (\alpha + \beta X)]^2$$

are given by

$$\alpha = \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2} \mu_X = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X$$
$$\beta = \frac{\sigma_{XY}}{\sigma_X^2} = \rho \frac{\sigma_Y}{\sigma_X}$$

where $\mu_Y = E[Y]$, $\mu_X = E[X]$, $\sigma_Y^2 = \text{Var}[Y]$, $\sigma_X^2 = \text{Var}[X]$, $\sigma_{XY} = \text{Cov}(X, Y)$, $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$.

4 Covariance matrix

4.1 Definition

Covariance matrix

$$C_{\mathbf{X}} := E [(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^T]$$

where the element (i, j)

$$C_{\mathbf{X}}(i, j) = E [(X_i - \mu_i)(X_j - \mu_j)]$$

is the covariance between X_i and X_j .

- Covariance matrix is nonnegative definite, i.e., for all \mathbf{x} we have

$$\mathbf{x}^T C_{\mathbf{X}} \mathbf{x} \geq 0$$

Hence

$$\det C_{\mathbf{X}} \geq 0.$$

- The covariance matrix is symmetric

$$C_{\mathbf{X}} = C_{\mathbf{X}}^T$$

4.2 2×2 Covariance Matrix

The covariance matrix of a bivariate random variable $\mathbf{X} = (X_1, X_2)^T$.

$$C_{\mathbf{X}} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where ρ is the coefficient of correlation of X_1 and X_2 , and $\sigma_1^2 = \text{Var}(X_1)$, $\sigma_2^2 = \text{Var}(X_2)$. $C_{\mathbf{X}}$ is invertible iff $\rho^2 \neq 1$. Then the inverse is

$$C_{\mathbf{X}}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

5 Discrete Random Variables

X is a (**discrete**) random variable that assumes values in \mathcal{X} and Y is a (**discrete**) random variable that assumes values in \mathcal{Y} .

Remark 5.1 These are measurable maps $X(\omega)$, $\omega \in \Omega$, from a basic probability space (Ω, \mathcal{F}, P) (= outcomes, a sigma field of subsets of Ω and probability measure P on \mathcal{F}), to \mathcal{X} . ■

\mathcal{X} and \mathcal{Y} are two discrete *state spaces*, whose generic elements are called *values* or *instantiations* and denoted by x_i and y_j , respectively.

$$\mathcal{X} = \{x_1, \dots, x_L\}, \mathcal{Y} = \{y_1, \dots, y_J\}.$$

$|\mathcal{X}|$ (:= the number of elements in \mathcal{X}) = $L \leq \infty$, $|\mathcal{Y}| = J \leq \infty$. Unless otherwise stated the alphabets considered here are finite.

5.1 Joint Probability Distributions

A two dimensional *joint (simultaneous) probability distribution* is a probability defined on $\mathcal{X} \times \mathcal{Y}$

$$p(x_i, y_j) := P(X = x_i, Y = y_j). \quad (5.1)$$

Hence $0 \leq p(x_i, y_j)$ and $\sum_{i=1}^L \sum_{j=1}^J p(x_i, y_j) = 1$.

Marginal distribution for X :

$$p(x_i) = \sum_{j=1}^J p(x_i, y_j). \quad (5.2)$$

Marginal distribution for Y :

$$p(y_j) = \sum_{i=1}^L p(x_i, y_j). \quad (5.3)$$

These notions can be extended to define the joint (simultaneous) probability distribution and the marginal distributions of n random variables.

5.2 Conditional Probability Distributions

The conditional probability for $X = x_i$ given $Y = y_j$ is

$$p(x_i | y_j) := \frac{p(x_i, y_j)}{p(y_j)}. \quad (5.4)$$

The conditional probability for $Y = y_j$ given $X = x_i$ is

$$p(y_j | x_i) := \frac{p(x_i, y_j)}{p(x_i)}. \quad (5.5)$$

Here we assume $p(y_j) > 0$ and $p(x_i) > 0$. If for example $p(x_i) = 0$, we can make the definition of $p(y_j | x_i)$ arbitrarily through $p(x_i) \cdot p(y_j | x_i) = p(x_i, y_j)$. In other words

$$p(y_j | x_i) = \frac{\text{prob. for the event } \{X = x_i, Y = y_j\}}{\text{prob. for the event } \{X = x_i\}}.$$

Hence

$$\sum_{i=1}^L p(x_i | y_j) = 1.$$

Next

$$P_X(A) := \sum_{x_i \in A} p(x_i) \quad (5.6)$$

is the probability of the event that X assumes a value in A , a subset of \mathcal{X} . From (5.6) one easily finds the complement rule

$$P_X(A^c) = 1 - P_X(A), \quad (5.7)$$

where A^c is the complement of A , i.e., those outcomes which do not lie in A . Also

$$P_X(A \cup B) = P_X(A) + P_X(B) - P_X(A \cap B), \quad (5.8)$$

is immediate.

5.3 Conditional Probability Given an Event

The conditional probability for $X = x_i$ given $X \in A$ is denoted by $P_X(x_i | A)$ and given by

$$P_X(x_i | A) = \begin{cases} \frac{P_X(x_i)}{P_X(A)} & \text{if } x_i \in A \\ 0 & \text{otherwise.} \end{cases} \quad (5.9)$$

5.4 Independence

X and Y are *independent* random variables if and only if

$$p(x_i, y_j) = p(x_i) \cdot p(y_j) \quad (5.10)$$

for *all pairs* (x_i, y_j) in $\mathcal{X} \times \mathcal{Y}$. In other words all events $\{X = x_i\}$ and $\{Y = y_j\}$ are to be independent. We say that X_1, X_2, \dots, X_n are **independent** random variables if and only if the joint distribution

$$p_{X_1, X_2, \dots, X_n}(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = P(X_1 = x_{i_1}, X_2 = x_{i_2}, \dots, X_n = x_{i_n}) \quad (5.11)$$

equals

$$p_{X_1, X_2, \dots, X_n}(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = p_{X_1}(x_{i_1}) \cdot p_{X_2}(x_{i_2}) \cdots p_{X_n}(x_{i_n}) \quad (5.12)$$

for every $x_{i_1}, x_{i_2}, \dots, x_{i_n} \in \mathcal{X}^n$. We are here assuming for simplicity that X_1, X_2, \dots, X_n take values in the same alphabet.

5.5 A Chain Rule

Let Z be a (discrete) random variable that assumes values in $\mathcal{Z} = \{z_k\}_{k=1}^K$. If $p(z_k) > 0$,

$$p(x_i, y_j | z_k) = \frac{p(x_i, y_j, z_k)}{p(z_k)}.$$

Then we obtain as an identity

$$p(x_i, y_j | z_k) = \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)} \cdot \frac{p(y_j, z_k)}{p(z_k)}$$

and again by definition of conditional probability the right hand side is equal to

$$p(x_i | y_j, z_k) \cdot p(y_j | z_k).$$

In other words,

$$p_{X,Y|Z}(x_i, y_j | z_k) = p(x_i | y_j, z_k) \cdot p(y_j | z_k). \quad (5.13)$$

5.6 Conditional Independence

The random variables X and Y are called *conditionally independent* given Z if

$$p(x_i, y_j | z_k) = p(x_i | z_k) \cdot p(y_j | z_k) \quad (5.14)$$

for all triples $(z_k, x_i, y_j) \in \mathcal{Z} \times \mathcal{X} \times \mathcal{Y}$ (cf. (5.13)).

6 Miscellaneous

6.1 A Marginalization Formula

Let Y be discrete and X be continuous, and let their joint distribution be

$$P(Y = k, X \leq x) = \int_{-\infty}^x P(Y = k | X = u) f_X(u) du.$$

Then

$$\begin{aligned} P(Y = k) &= \int_{-\infty}^{\infty} \frac{\partial}{\partial u} P(Y = k, X \leq u) du \\ &= \int_{-\infty}^{\infty} P(Y = k | X = x) f_X(x) dx. \end{aligned}$$

6.2 Factorial Moments

X is an integer-valued discrete R.V.,

$$\begin{aligned} \mu_{[r]} &\stackrel{def}{=} E[X(X-1)\cdots(X-r+1)] = \\ &= \sum_{x:\text{integer}} (x(x-1)\cdots(x-r+1)) f_X(x). \end{aligned}$$

is called the r :th factorial moment.

6.3 Binomial Moments

X is an integer-valued discrete R.V..

$$E\binom{X}{r} = E[X(X-1)\cdots(X-r+1)]/r!$$

is called the binomial moment.

7 Transforms

7.1 Probability Generating Function

7.1.1 Definition

Let X have values $k = 0, 1, 2, \dots$.

$$g_X(t) = E(t^X) = \sum_{k=0}^{\infty} t^k f_X(k)$$

is called the probability generating function.

7.1.2 Prob. Gen. Fnct: Properties

•

$$\begin{aligned} \frac{d}{dt} g_X(1) &= \sum_{k=1}^{\infty} k t^{k-1} f_X(k) \Big|_{t=1} \\ &= E[X] \end{aligned}$$

•

$$\frac{d^n g_X(0)}{n!} = P(X = n)$$

•

$$\mu_{[r]} = E[X(X-1)\cdots(X-r+1)] = \frac{d^r}{dt^r} g_X(1)$$

7.1.3 Prob. Gen. Fnct: Properties

$Z = X + Y$, X and Y non negative integer valued, independent,

•

$$\begin{aligned} g_Z(t) &= E(t^Z) = \\ E(t^{X+Y}) &= E(t^X) \cdot E(t^Y) = g_X(t) \cdot g_Y(t). \end{aligned}$$

7.1.4 Prob. Gen. Fnct: Examples

- $X \in \text{Be}(p)$

$$g_X(t) = 1 - p + pt.$$

- $Y \in \text{Bin}(n, p)$

$$g_Y(t) = (1 - p + pt)^n$$

- $Z \in \text{Po}(\lambda)$

$$g_Z(t) = e^{\lambda(t-1)}$$

7.2 Moment Generating Functions

7.2.1 Definition

Moment generating function, for some $h > 0$,

$$\psi_X(t) \stackrel{\text{def}}{=} E[e^{tX}], |t| < h.$$

$$\psi_X(t) = \begin{cases} \sum_{x_i} e^{tx_i} f_X(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x) dx & X \text{ continuous} \end{cases}$$

7.2.2 Moment Gen. Fnctn: Properties

-
-

$$\frac{d}{ds} \psi_X(0) = E[X]$$

$$\psi_X(0) = 1$$

$$\frac{d^k}{ds^k} \psi_X(0) = E[X^k].$$

$S_n = X_1 + X_2 + \dots + X_n$, X_i independent.

$$\psi_{S_n}(t) = E(e^{tS_n}) =$$

$$E(e^{t(X_1+X_2+\dots+X_n)}) = E(e^{tX_1} e^{tX_2} \dots e^{tX_n}) =$$

$$E(e^{tX_1}) E(e^{tX_2}) \dots E(e^{tX_n}) = \psi_{X_1}(t) \cdot \psi_{X_2}(t) \dots \psi_{X_n}(t)$$

X_i I.I.D.,

$$\psi_{S_n}(s) = (\psi_X(t))^n.$$

7.2.3 Moment Gen. Fnctn: Examples

- $X \in N(\mu, \sigma^2)$

$$\psi_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

- $Y \in \text{Exp}(a)$

$$\psi_Y(s) = \frac{1}{1 - at}, \quad at < 1.$$

7.2.4 Characteristic function

Characteristic function

$$\varphi_X(t) = E[e^{itX}]$$

exists for all t .

7.3 Moment generating function, characteristic function of a vector random variable

Moment generating function of \mathbf{X} ($n \times 1$ vector) is defined as

$$\psi_{\mathbf{X}}(\mathbf{t}) \stackrel{\text{def}}{=} Ee^{\mathbf{t}^T \mathbf{X}} = Ee^{t_1 X_1 + t_2 X_2 + \dots + t_n X_n}$$

Characteristic function of \mathbf{X} is defined as

$$\varphi_{\mathbf{X}}(\mathbf{t}) \stackrel{\text{def}}{=} Ee^{i\mathbf{t}^T \mathbf{X}} = Ee^{i(t_1 X_1 + t_2 X_2 + \dots + t_n X_n)}$$

8 Multivariate normal distribution

An $n \times 1$ random vector \mathbf{X} has a normal distribution iff for every $n \times 1$ -vector \mathbf{a} the one-dimensional random vector $\mathbf{a}^T \mathbf{X}$ has a normal distribution.

■

When vector $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ has a multivariate normal distribution we write

$$\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{C}). \quad (8.1)$$

The moment generating function is

$$\psi_{\mathbf{X}}(\mathbf{s}) = e^{\mathbf{s}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{s}^T \mathbf{C} \mathbf{s}} \quad (8.2)$$

The characteristic function of $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{C})$ is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = Ee^{it^T \mathbf{X}} = e^{it^T \boldsymbol{\mu} - \frac{1}{2} \mathbf{t}^T \mathbf{C} \mathbf{t}}. \quad (8.3)$$

Let $\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{C})$ and

$$\mathbf{Y} = \mathbf{a} + B\mathbf{X}.$$

for an arbitrary $m \times n$ -matrix B and arbitrary $m \times 1$ vector \mathbf{a} . Then

$$\mathbf{Y} \in N(\mathbf{a} + B\boldsymbol{\mu}, B\mathbf{C}B^T). \quad (8.4)$$

8.1 χ^2 -distribution for quadratic form of normal r.v.

$\mathbf{X} \in N(\boldsymbol{\mu}, \mathbf{C})$, \mathbf{C} is invertible. Then

$$(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \in \chi^2(n),$$

where n is the dimension of \mathbf{X} .

8.2 Four product rule

$(X_1, X_2, X_3, X_4)^T \in N(\underline{0}, \mathbf{C})$. Then

$$E[X_1 X_2 X_3 X_4] = E[X_1 X_2] \cdot E[X_3 X_4] + E[X_1 X_3] \cdot E[X_2 X_4] + E[X_1 X_4] \cdot E[X_2 X_3]$$

8.3 Conditional distributions for bivariate normal random variables

$(X, Y)^T \in N(\underline{\boldsymbol{\mu}}, \mathbf{C})$. The *conditional* distribution for Y given $X = x$ is normal

$$N\left(\mu_Y + \rho \cdot \frac{\sigma_Y}{\sigma_X} (x - \mu_X), \sigma_Y^2 (1 - \rho^2)\right),$$

where $\mu_Y = E(Y)$, $\mu_X = E(X)$, $\sigma_Y = \sqrt{\text{Var}(Y)}$, $\sigma_X = \sqrt{\text{Var}(X)}$ and

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}.$$

Z_1 och Z_2 are independent $N(0, 1)$.

$$\underline{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}.$$

If

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \underline{\mu} + B \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

then

$$\begin{pmatrix} X \\ Y \end{pmatrix} \in N(\underline{\mu}, \mathbf{C})$$

with

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

9 Poisson process

$X(t)$ = number of occurrences of some event in in $(0, t]$.

Definition 9.1 $\{X(t) \mid t \geq 0\}$ is a Poisson process with parameter $\lambda > 0$, if

- 1) $X(0) = 0$.
- 2) The increments $X(t_k) - X(t_{k-1})$ are independent stochastic variables $1 \leq k \leq n$, $0 \leq t_0 \leq t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq t_n$ and all n .
- 3) $X(t) - X(s) \in \text{Po}(\lambda(t-s))$, $0 \leq s < t$.

■

T_k = the time of occurrence of the k th event. $T_0 = 0$. We have

$$\{T_k \leq t\} = \{X(t) \geq k\}$$

$$\tau_k = T_k - T_{k-1},$$

is the k th interoccurrence time. $\tau_1, \tau_2, \dots, \tau_k, \dots$ are independent and identically distributed, $\tau_i \in \text{Exp}\left(\frac{1}{\lambda}\right)$.

10 Convergence

10.1 Definitions

- We say that

$$X_n \xrightarrow{P} X, \quad \text{as } n \rightarrow \infty$$

if for all $\epsilon > 0$

$$P(|X_n - X| > \epsilon) \rightarrow 0, \quad \text{as } n \rightarrow \infty$$

■

- We say that

$$X_n \xrightarrow{q} X$$

if

$$E|X_n - X|^2 \rightarrow 0, \quad \text{as } n \rightarrow \infty$$

■

- We say that

$$X_n \xrightarrow{d} X, \quad \text{as } n \rightarrow \infty$$

if

$$F_{X_n}(x) \rightarrow F_X(x), \quad \text{as } n \rightarrow \infty$$

for all x , where $F_X(x)$ is continuous.

■

10.2 Relations between convergences

$$X_n \xrightarrow{q} X \Rightarrow X_n \xrightarrow{P} X$$

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$$

as $n \rightarrow \infty$. If c is a constant,

$$X_n \xrightarrow{P} c \Leftrightarrow X_n \xrightarrow{d} c$$

as $n \rightarrow \infty$.

If $\varphi_{X_n}(t)$ are the characteristic functions of X_n , then

$$X_n \xrightarrow{d} X \Rightarrow \varphi_{X_n}(t) \rightarrow \varphi_X(t)$$

If $\varphi_X(t)$ is a characteristic function continuous at $t = 0$, then

$$\varphi_{X_n}(t) \rightarrow \varphi_X(t) \Rightarrow X_n \xrightarrow{d} X$$

10.3 Law of Large Numbers

X_1, X_2, \dots are independent, identically distributed (i.i.d.) random variables with finite expectation μ . We set

$$S_n = X_1 + X_2 + \dots + X_n, \quad n \geq 1.$$

Then

$$\frac{S_n}{n} \xrightarrow{P} \mu, \quad \text{as } n \rightarrow \infty.$$

10.4 Central Limit Theorem

X_1, X_2, \dots are independent, identically distributed (i.i.d.) random variables with finite expectation μ and finite variance σ^2 . We set

$$S_n = X_1 + X_2 + \dots + X_n, \quad n \geq 1.$$

Then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1), \quad \text{as } n \rightarrow \infty.$$

11 Series Expansions and Integrals

11.1 Exponential Function

•

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad -\infty < x < \infty.$$

•

$$c_n \rightarrow c \Rightarrow \left(1 + \frac{c_n}{n}\right)^n \rightarrow e^c.$$

11.2 Geometric Series

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k, \quad |x| < 1.$$

$$\frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} kx^{k-1}, \quad |x| < 1.$$

$$\sum_{k=0}^n x^k = \frac{1-x^{n+1}}{1-x}, \quad x \neq 1.$$

11.3 Logarithm function

$$-\ln(1-x) = \sum_{k=1}^{\infty} \frac{x^k}{k}, \quad -1 \leq x < 1.$$

11.4 Euler Gamma Function

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx, \quad t > 0$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma(n) = (n-1)! \quad n \text{ is a nonnegative integer.}$$

$$\int_0^{\infty} x^t e^{-\lambda x} dx = \frac{\Gamma(t+1)}{\lambda^{t+1}}, \quad \lambda > 0, t > -1$$

11.5 A formula (with a probabilistic proof)

$$\int_t^{\infty} \frac{1}{\Gamma(k)} \lambda^k x^{k-1} e^{-\lambda x} dx = \sum_{j=0}^{k-1} e^{-\lambda t} \frac{(\lambda t)^j}{j!}$$