Formulas for probability theory SF2940 (23 pages)

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

6 september 2011

- Selected formulae of probability
- Bivariate probability
- Conditional expectation w.r.t a Sigma field
- Transforms
- Multivariate normal distribution
- Stochastic processes
- Gaussian processes
- Poisson process
- Convergence
- Series Expansions and Integrals

1 Probability

1.1 Two inequalities

- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(A \cup B) \leq P(A) + P(B)$ (Boole's inequality).

1.2 Change of variable in a probability density

Let $\mathbf{X} = (X_1, X_2, \ldots, X_m)^T$ have the probability density $f_{\mathbf{X}}(x_1, x_2, \ldots, x_m)$. Define a new random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)^T$ by

$$Y_i = g_i(X_1, \ldots, X_m), \quad i = 1, 2, \ldots, m,$$

where g_i are continuously differentiable and $(g_1, g_2, \ldots, g_m)$ is invertible (in a domain) with

$$X_i = h_i(Y_1, \ldots, Y_m), \quad i = 1, 2, \ldots, m,$$

where h_i are continuously differentiable. Then the density of $\mathbf{Y}$ is (in the domain of invertibility)

$$f_{\mathbf{Y}}(y_1, \ldots, y_m) = f_{\mathbf{X}}(h_1(y_1, y_2, \ldots, y_m), \ldots, h_m(y_1, y_2, \ldots, y_m)) \mid J \mid,$$

where J is the Jacobian determinant

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \ldots & \frac{\partial x_1}{\partial y_m} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \ldots & \frac{\partial x_2}{\partial y_m} \\ \vdots & \vdots & \ldots & \vdots \\ \frac{\partial x_m}{\partial y_1} & \frac{\partial x_m}{\partial y_2} & \ldots & \frac{\partial x_m}{\partial y_m} \end{vmatrix}.$$

Example 1.1 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

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

■

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

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

  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:

- $X \mid Y = y$,

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

  if $f_Y(y) > 0$.

- $Y \mid X = x$

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

  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 approximation[1]:

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

[1] P.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.

2.2 Mean and variance

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

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

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

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

respectively. We have

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

The *function of a random variable* $g(X)$, **the law of the unconscious statistician**,

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

2.3 Chebyshev's inequality

$$P\left(\mid X - E(X) \mid > \varepsilon\right) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

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 $y \longmapsto E(X|Y = y)$, as a function of Y.

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

2.5 Covariance

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

We have

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

2.6 Coefficient of correlation

Coefficient of correlation between X and Y is defined as

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

3 Best linear prediction

α and β that minimize

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

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

4 Conditional Expectation w.r.t to a Sigma-Field

a and b are real numbers, $E[| Y |] < \infty$, $E[| Z |] < \infty$, $E[| X |] < \infty$ and $\mathcal{H}$, $\mathcal{G}$ $\mathcal{F}$ are sigma fields, $\mathcal{G} \subset \mathcal{F}$, $\mathcal{H} \subset \mathcal{F}$.

1. **Linearity**:
$$E[aX + bY \mid \mathcal{G}] = aE[X \mid \mathcal{G}] + bE[Y \mid \mathcal{G}]$$

2. **Double expectation** :
$$E[E[Y \mid \mathcal{G}]] = E[Y]$$

3. **Taking out what is known**: If Z is $\mathcal{G}$ -measurable, and $E[| ZY |] < \infty$
$$E[ZY \mid \mathcal{G}] = ZE[Y \mid \mathcal{G}]$$

4. **An independent condition drops out**: If Y is independent of $\mathcal{G}$,
$$E[Y \mid \mathcal{G}] = E[Y]$$

5. **Tower Property** : If $\mathcal{H} \subset \mathcal{G}$,
$$E[E[Y \mid \mathcal{G}] \mid \mathcal{H}] = E[Y \mid \mathcal{H}]$$

6. **Positivity**: If $Y \geq 0$,
$$E[Y \mid \mathcal{G}] \geq 0.$$

5 Covariance matrix

5.1 Definition

Covariance matrix

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

where the entry in position (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 $\mathbf{x} \neq \mathbf{0}$ we have

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

Hence

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

- The covariance matrix is symmetric

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

5.2 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 $\sigma_1^2 = \mathrm{Var}\,(X_1)$, $\sigma_2^2 = \mathrm{Var}\,(X_2)$. $C_{\mathbf{X}}$ is invertible iff $\rho^2 \neq 1$, then the inverse is

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

6 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 6.1 These are measurable maps $X(\omega)$, $\omega \in \Omega$, from a basic probability space $(\Omega, \mathcal{F}, P)$ (= outcomes, a sigma field of 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.

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

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

6.1 Joint Probability Distributions

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

$$p(x_i, y_j) := P(X = x_i, Y = y_j). \tag{6.1}$$

Hence $0 \leq p(x_i, y_j)$ and $\sum_{i=1}^{L} \sum_{j=1}^{L} p(x_i, y_j) = 1$.
Marginal distribution for X:

$$p(x_i) = \sum_{j=1}^{J} p(x_i, y_j). \tag{6.2}$$

Marginal distribution for Y:

$$p(y_j) = \sum_{i=1}^{L} p(x_i, y_j). \tag{6.3}$$

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

6.2 Conditional Probability Distributions

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

$$p(x_i \mid y_j) := \frac{p(x_i, y_j)}{p(y_j)}. \tag{6.4}$$

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

$$p(y_j \mid x_i) := \frac{p(x_i, y_j)}{p(x_i)}. \tag{6.5}$$

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 $p(y_j \mid x_i)$ arbitrarily through $p(x_i) \cdot p(y_j \mid x_i) = p(x_i, y_j)$. In other words

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

Hence

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

Next

$$P_X(A) := \sum_{x_i \in A} p(x_i) \tag{6.6}$$

is the probability of the event that X assumes a value in A, a subset of $\mathcal{X}$. From (6.6) one easily finds the complement rule

$$P_X\left(A^{\mathrm{c}}\right) = 1 - P_X(A), \tag{6.7}$$

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

$$P_X\left(A \cup B\right) = P_X(A) + P_X(B) - P_X(A \cap B), \tag{6.8}$$

is immediate.

6.3 Conditional Probability Given an Event

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

$$P_X\left(x_i \mid A\right) = \begin{cases} \frac{P_X(x_i)}{P_X(A)} & \text{if } x_i \in A \\ 0 & \text{otherwise.} \end{cases} \tag{6.9}$$

6.4 Independence

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

$$p(x_i, y_j) = p(x_i) \cdot p(y_j) \tag{6.10}$$

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 $X_1, X_2, \ldots, X_n$ are **independent** random variables if and only if the joint distribution

$$p_{X_1,X_2,\ldots,X_n}(x_{i_1}, x_{i_2} \ldots, x_{i_n}) = P\left(X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_n = x_{i_n}\right) \tag{6.11}$$

equals

$$p_{X_1,X_2,\ldots,X_n}(x_{i_1}, x_{i_2} \ldots, x_{i_n}) = p_{X_1}(x_{i_1}) \cdot p_{X_2}(x_{i_2}) \cdots p_{X_n}(x_{i_n}) \tag{6.12}$$

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

6.5 A Chain Rule

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

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

Then we obtain as an identity

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

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

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

In other words,

$$p_{X,Y|Z}(x_i, y_j \mid z_k) = p(x_i \mid y_j, z_k) \cdot p(y_j \mid z_k). \tag{6.13}$$

6.6 Conditional Independence

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

$$p(x_i, y_j | z_k) = p(x_i | z_k) \cdot p(y_j | z_k) \tag{6.14}$$

for all triples $(z_k, x_i, y_j) \in \mathcal{Z} \times \mathcal{X} \times \mathcal{Y}$ (cf. (6.13)).

7 Miscellaneous

7.1 A Marginalization Formula

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

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

Then

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

7.2 Factorial Moments

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

$$\mu_{[r]} \stackrel{def}{=} E\left[X(X-1)\cdots(X-r+1)\right] =$$

$$= \sum_{x:\text{integer}} \left(x(x-1)\cdots(x-r+1)\right) f_X(x).$$

is called the r:th factorial moment.

7.3 Binomial Moments

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

$$E\begin{pmatrix} X \\ r \end{pmatrix} = E\left[X(X-1)\cdots(X-r+1)\right]/r!$$

is called the binomial moment.

8 Transforms

8.1 Probability Generating Function

8.1.1 Definition

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

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

is called the probability generating function.

8.1.2 Prob. Gen. Fnct: Properties

- $$\frac{d}{dt} g_X(1) = \sum_{k=1}^{\infty} k t^{k-1} f_X(k) \mid_{t=1}$$
  $$= E[X]$$

- $$\mu_{[r]} = E\left[X(X-1)\cdots(X-r+1)\right] = \frac{d^r}{dt^r} g_X(1)$$
$$\mathrm{Var}[X] = \frac{d^2}{dt^2} g_X(1) + \frac{d}{dt} g_X(1) - \left(\frac{d}{dt} g_X(1)\right)^2$$

8.1.3 Prob. Gen. Fnct: Properties

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

- $$g_Z(t) = E\left(t^Z\right) =$$
$$E\left(t^{X+Y}\right) = E\left(t^X\right) \cdot E\left(t^Y\right) = g_X(t) \cdot g_Y(t).$$

8.1.4 Prob. Gen. Fnct: Examples

- $X \in \mathrm{Be}(p)$
$$g_X(t) = 1 - p + pt.$$
- $Y \in \mathrm{Bin}(n, p)$
$$g_Y(t) = (1 - p + pt)^n$$
- $Z \in \mathrm{Po}(\lambda)$
$$g_Z(t) = e^{\lambda \cdot (t-1)}$$

8.1.5 Sum of a Random Number of Random Variables

X_i., $i = 1, \ldots, n$ I.I.D. non negative integer valued, and N non negative integer valued and independent of the X_is.

$$S_N = \sum_{i=1}^{N} X_i.$$

Then the probability generating function of S_N is

$$g_{S_N}(t) = g_N\left(g_X(t)\right).$$

8.2 Moment Generating Functions

8.2.1 Definition

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

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

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

8.2.2 Moment Gen. Fnctn: Properties

- $$\frac{d}{dt}\psi_X(0) = E[X]$$
- $$\psi_X(0) = 1$$
  $$\frac{d^k}{dt^k}\psi_X(0) = E\left[X^k\right].$$

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

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

$$E\left(e^{t(X_1+X_2+\ldots+X_n)}\right) = E\left(e^{tX_1}e^{tX_2}\cdots e^{tX_n}\right) =$$

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

X_i I.I.D.,

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

8.2.3 Moment Gen. Fnctn: Examples

- $X \in \mathrm{N}(\mu, \sigma^2)$
  $$\psi_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$
- $Y \in \mathrm{Exp}(a)$
  $$\psi_Y(s) = \frac{1}{1-at}, \quad at < 1.$$

8.2.4 Characteristic function

Characteristic function

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

exists for all t.

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

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

$$\psi_{\mathbf{X}}(\mathbf{t}) \stackrel{\text{def}}{=} Ee^{\mathbf{t}^T\mathbf{X}} = Ee^{t_1X_1+t_2X_2+\cdots+t_nX_n}$$

Characteristic function of $\mathbf{X}$ is defined as

$$\varphi_{\mathbf{X}}(\mathbf{t}) \stackrel{\text{def}}{=} Ee^{i\mathbf{t}^T\mathbf{X}} = Ee^{i(t_1X_1+t_2X_2+\cdots+t_nX_n)}$$

9 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 $a^T\mathbf{X}$ has a normal distribution.
■

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

$$\mathbf{X} \in N(\mu, \mathbf{C}). \tag{9.1}$$

The moment generating function is

$$\psi_{\mathbf{X}}(\mathbf{s}) = e^{\mathbf{s}^T\mu+\frac{1}{2}\mathbf{s}^T\mathbf{C}\mathbf{s}} \tag{9.2}$$

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

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

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

$$\mathbf{Y} \in N\left(\mathbf{a} + B\mu, B\mathbf{C}B^T\right). \tag{9.3}$$

9.1 Four product rule

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

$E[X_1X_2X_3X_4] = E[X_1X_2]{\cdot}E[X_3X_4]{+}E[X_1X_3]{\cdot}E[X_2X_4]{+}E[X_1X_4]{\cdot}E[X_2X_3]$

9.2 Conditional distributions for bivariate normal random variables

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

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

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

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

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

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

If

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

then

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

with

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

10 Stochastic Processes

A stochastic process $\mathbf{X} = \{X(t) \mid t \in T\}$. The *mean function* $\mu_{\mathbf{X}}(t)$ of the process is

$$\mu_{\mathbf{X}}(t) \stackrel{def}{=} E(X(t))$$

and the *autocorrelation function* is

$$R_{\mathbf{X}}(t,s) \stackrel{def}{=} E\left(X(t) \cdot X(s)\right).$$

The *autocovariance* function is

$$\mathrm{Cov}_{\mathbf{X}}(t,s) \stackrel{def}{=} E\left((X(t) - \mu(t)) \cdot (X(s) - \mu(s))\right)$$

and we have

$$\mathrm{Cov}_{\mathbf{X}}(t,s)(t,s) = R_{\mathbf{X}}(t,s) - \mu_{\mathbf{X}}(t)\mu_{\mathbf{X}}(s).$$

A stochastic process $\mathbf{X} = \{X(t) \mid t \in T =]-\infty, \infty[\}$ is called **(weakly) stationary** if

1. The mean function $\mu_{\mathbf{X}}(t)$ is a constant function of t, $\mu_{\mathbf{X}}(t) = \mu$.

2. The autocorrelation function $R_{\mathbf{X}}(t,s)$ is a function of $(t-s)$, so that

$$R_{\mathbf{X}}(t,s) = R_{\mathbf{X}}(h) = R_{\mathbf{X}}(-h), \quad h = (t-s).$$

10.1 Mean Square Integrals

$$\sum_{i=1}^{n} X(t_i)(t_i - t_{i-1}) \xrightarrow{2} \int_a^b X(t)dt, \tag{10.4}$$

where $a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b$ and $\max_i |t_i - t_{i-1}| \to 0$ as $n \to \infty$.

The mean square integral $\int_a^b X(t)dt$ of $\{X(t)|t \in T\}$ exists over $[a,b] \subseteq T$ if and only if the double integral

$$\int_a^b \int_a^b E\left[X(t)X(u)\right] dtdu$$

exists as an integral in Riemann's sense. We have also

$$E\left[\int_a^b X(t)dt\right] = \int_a^b \mu_{\mathbf{X}}(t)dt \tag{10.5}$$

and

$$Var\left[\int_a^b X(t)dt\right] = \int_a^b \int_a^b \text{Cov}_{\mathbf{X}}(t,u)dtdu. \tag{10.6}$$

$\mathbf{X} = \{X(t)|t \in T\}$ is a stochastic process. Then the process is *mean square continuous* if, when $t + \tau \in T$,

$$E\left[(X(t+\tau) - X(t))^2\right] \to 0$$

as $\tau \to 0$.

10.2 Gaussian stochastic processes

A stochastic process $\mathbf{X} = \{X(t) \mid -\infty \leq t \leq \infty\}$ is *Gaussian*, if every random n-vector $(X(t_1), X(t_2), \cdots, X(t_n))$ is a multivariate normal vector. ■

10.3 Wiener process

A Wiener process $\mathbf{W}$ is a Gaussian process such that $W(0) = 0$, $\mu\mathbf{W}(t) = 0$ for all $t \geq 0$ and

$$E\left[W(t) \cdot W(s)\right] = \min(t, s).$$

1) $W(0) = 0$.

2) The sample paths $t \mapsto W(t)$ are almost surely continuous.

3) $\{W(t) \mid t \geq 0\}$ has stationary and independent increments.

4) $W(t) - W(s) \in N(0, t-s)$ for $t > s$.

10.4 Wiener Integrals

$f(t)$ is a function such that $\int_a^b f^2(t)dt < \infty$, where $-\infty \leq a < b \leq +\infty$. The mean square integral with respect to the Wiener process or the **Wiener integral** $\int_a^b f(t)dW(t)$ is the mean square limit

$$\sum_{i=1}^{n} f(t_{i-1})\left(W(t_i) - W(t_{i-1})\right) \xrightarrow{2} \int_a^b f(t)dW(t), \tag{10.7}$$

$a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b$ and $\max_i |t_i - t_{i-1}| \to 0$ as $n \to \infty$.

- $$E\left[\int_a^b f(t)dW(t)\right] = 0. \tag{10.8}$$

- $$\text{Var}\left[\int_a^b f(t)dW(t)\right] = \int_a^b f^2(t)dt \tag{10.9}$$

- $$\int_a^b f(t)dW(t) \in N\left(0, \int_a^b f^2(t)dt\right). \tag{10.10}$$

- If $\int_a^b f^2(t)dt < \infty$ and $\int_a^b g^2(t)dt < \infty$,
$$E\left[\int_a^b f(t)dW(t) \int_a^b g(t)dW(t)\right] = \int_a^b f(t)g(t)dt. \tag{10.11}$$

- $$Y(t) = \int_0^t h(s)dW(s).$$
$$E\left[Y(t) \cdot Y(s)\right] = \int_0^{\min(t,s)} h^2(u)du. \tag{10.12}$$

11 Poisson process

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

Definition 11.1 $\{N(t) \mid t \geq 0\}$ is a Poisson process with parameter $\lambda > 0$, if

1) $N(0) = 0$.

2) The increments $N(t_k) - N(t_{k-1})$ are independent stochastic variables $1 \leq k \leq n$, $0 \leq t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_{n-1} \leq t_n$ and all n.

3) $N(t) - N(s) \in \text{Po}(\lambda(t-s))$, $\quad 0 \leq s < t$.

■

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

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

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

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

12 Convergence

12.1 Definitions

We say that

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

if for all $\epsilon > 0$

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

■

We say that

$$X_n \xrightarrow{q} X$$

if

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

■

We say that

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

if

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

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

12.2 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 $n \to \infty$. If c is a constant,

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

as $n \to \infty$.

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

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

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

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

12.3 Law of Large Numbers

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

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

Then

$$\frac{S_n}{n} \xrightarrow{P} \mu, \qquad \text{as } n \to \infty.$$

12.4 Borel-Cantelli lemmas

$$E = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

i.e.,

$$E = \{ \ A_k \text{ occurs infinitely often } \}$$

$$H = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

Lemma 12.1 Let $\{A_k\}_{k\geq 1}$ be arbitrary events. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then it holds that $P(E) = P(A_n \text{ i.o}) = 0$, ie., with probability one finitely many of A_n occur.

■

Lemma 12.2 Let $\{A_k\}_{k\geq 1}$ be independent events. If $\sum_{n=1}^{\infty} P(A_n) = \infty$, then it holds that $P(E) = P(A_n \text{ i.o}) = 1$, ie., with probability one infinitely many of A_n occur.

■

12.5 Central Limit Theorem

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

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

Then

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

13 Series Expansions and Integrals

13.1 Exponential Function

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

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

13.2 Geometric Series

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

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

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

13.3 Logarithm function

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

13.4 Euler Gamma Function

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

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

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

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

13.5 A formula (with a probabilistic proof)

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