NMAI059 Probability and statistics 1
Class 9

Robert Šámal

Overview

Continuous random vectors

Covariance and correlation

Inequalities

Limit theorems – approximation

What we know

(for discrete $P(X \le x \,\&\, Y \le y) = \sum_{s \le x} \sum_{t \le y} P(X = s \,\&\, Y = t)$)

- joint cdf

$$F_{X,Y}(x,y) = P(X \le x \,\&\, Y \le y).$$

- joint pdf: $f_{X,Y} \ge 0$ such that

$$\boxed{F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)\,dt\,ds.}$$

- important example: multivariate normal distribution

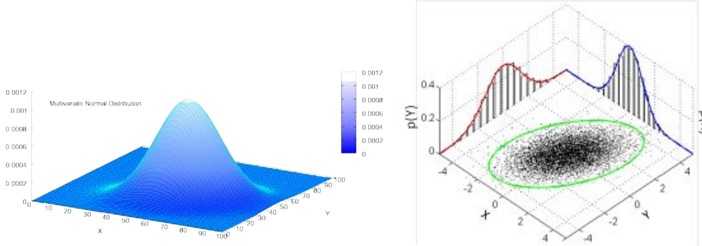
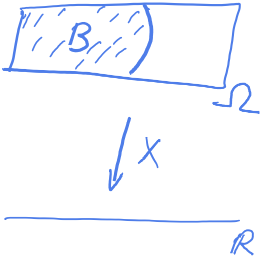


Image by Wikipedia editors Piotrg and Bscan.

Conditioning

$X|B :=$ r.v. X restricted to B



Definition (restricting a r.v. to a subset)

X is a r.v. on $(\Omega, \mathcal{F}, P)$, $B \in \mathcal{F}$, s.t. $P(B) > 0$.

$$F_{X|B}(x) := P(X \le x \mid B)$$

$f_{X|B}$ is the corresponding pdf. $= (F_{X|B})'$

- if $B = \{X \in S\}$, then

$$f_{X|B}(x) = \begin{cases} \frac{f_X(x)}{P(X \in S)} & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases} \quad (x \notin S)$$

$$P(B) = P(X \in S)$$

Total cdf & pdf

Theorem (total cdf, total pdf)

Let X be a continuous r.v., let $B_1, B_2, \ldots$ be a partition of Ω. Then

$$F_X(x) = \sum_i P(B_i) F_{X|B_i}(x) \quad \textit{and}$$

different.

$$f_X(x) = \sum_i P(B_i) f_{X|B_i}(x).$$

Proof: law of total probability.

$$= P(X \le x) = \sum_i P(B_i) \underbrace{P(X \le x \mid B_i)}_{F_{X|B}(x)}$$

X -- running time of an algo

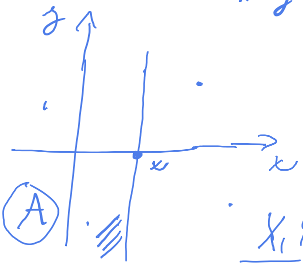
$B_1, B_2, \ldots$ different branches

Marginal pdf

discrete r. vector $p_X(x) = \sum_y p_{X,Y}(x,y)$

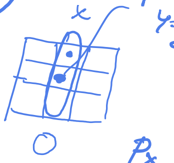
Theorem

marginal pdf joint pdf



$$f_X(x) = \int_{y\in\mathbb{R}} f_{X,Y}(x,y)dy$$

$$f_Y(y) = \int_{x\in\mathbb{R}} f_{X,Y}(x,y)dx$$



$P(X=x, Y=y)$

$\sum$ p_X

X, Y

Pf $F_X(x) = P(X \le x) = P((X,Y)\in A) = \int_A f_{X,Y}(s,t)$

$A = \{(s,t)\in\mathbb{R}^2 : s < x\}$

$= \int_{-\infty}^{x} \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(s,t)\,dt}_{g(s)}\,ds$

$F_X(x) = \int_{-\infty}^{x} g(s)\,ds$

$\Rightarrow g$ is the pdf of X

Conditional pdf

$$P(X=x \mid Y=y) = \frac{P(X=x \,\&\, Y=y)}{P(Y=y)}$$

Definition

For continuous r.v. X, Y we define their conditional pdf by

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

y

when $f_Y(y) > 0$, otherwise we do not define it.

- recall that $\boxed{f_Y(y) = \int_{x\in\mathbb{R}} f_{X,Y}(x,y)dx}$
- for a fixed y the function $x \mapsto f_{X|Y}(x|y)$ is a pdf

$$\int f_{X|Y}(x|y)\,dx = 1$$

$$\| \qquad \frac{\int f_{X,Y}(x,y)\,dx}{\boxed{f_Y(y)}}$$

Conditional, joint and marginal pdf

Theorem

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

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

$$f_X(x) = \int_{-\infty}^{\infty} \underline{f_{X,Y}(x,y)}\,dy$$

Sum of continuous r.v.

Theorem

Let X, Y be independent random variables. Then $Z = X + Y$ is also a continuous r.v. and its pdf is a convolution of f_X, f_Y. That is,

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx.}$$

for discrete r.v.

$$P_Z(z) = \sum_x P_X(x) \cdot P_Y(z-x)$$

Proof $f_{Z/X}(z/x) = f_Y(z-x)$

$Z/X = x$ the same as $Y + x$

$y + x = z \Leftrightarrow y = z - x$

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z/X}(z/x) \cdot f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)$$

Example of a convolution

$X, Y \sim N(0,1)$ independent $\quad f_X = f_Y = \varphi \quad \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$

$Z = X + Y$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \qquad (z^2 - z2x + x^2)$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}}\, dx$$

$$= \frac{1}{2\pi} e^{-z^2/2} \int_{-\infty}^{\infty} \underbrace{e^{-\frac{x^2}{2}} e^{+zx - \frac{x^2}{2}}} = \frac{1}{2\pi} e^{-\frac{z^2}{2} + \frac{z^2}{4}} \int_{-\infty}^{\infty} e^{-(x-\frac{z}{2})^2}\, dx$$

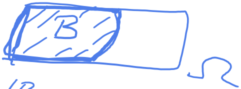
$$e^{-x^2 + zx} = e^{-(x - \frac{z}{2})^2 + \frac{z^2}{4}}$$

$$\int_{-\infty}^{\infty} e^{-x^2} = \sqrt{\pi}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2}} e^{-\frac{z^2}{4}}$$

$$\boxed{Z \sim N(0, 2)}$$

Conditional density and expectation



- $\mathbb{E}(X \mid B) := \int_{-\infty}^{\infty} x \cdot f_{X|B}(x)dx$
- $\mathbb{E}(g(X)|B) = \int_{-\infty}^{\infty} g(x) f_{X|B}(x)dx$

X/B is a r.v.

LOTUS

Theorem (total expectation)

Let X be a continuous r.v. If $B_1, B_2, \ldots$ is a partition of Ω, then

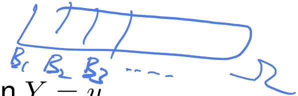
$$\boxed{\mathbb{E}(X) = \sum_i P(B_i)\mathbb{E}(X \mid B_i).}$$

if $\mathbb{E}(X|B_i)$ are defined

Proof: by total pdf.

$$\int_{-\infty}^{\infty} x f_X(x) = \int_{-\infty}^{\infty} x \sum_i P(B_i) f_{X|B_i}(x) = \sum_i P(B_i) \underbrace{\int_{-\infty}^{\infty} x \cdot f_{X|B_i}(x)}_{\mathbb{E}(X|B_i)}$$

Conditional pdf and expectation



- $f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$ is a pdf of X, given $Y = y$
- $\mathbb{E}\big(X \mid Y = y\big) := \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x,y)dx$ is the expectation of this r.v.
- $\mathbb{E}\big(g(X)\big|Y = y\big) = \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y}(x,y)dx$
- An analogy of the law of total expectation:

$$\mathbb{E}(X) = \sum P(B_i)\, \mathbb{E}(X \mid B_i)$$

$$\mathbb{E}\big(X\big) = \int_{-\infty}^{\infty} \underbrace{\mathbb{E}\big(X \mid Y = y\big)}_{g(y)} f_Y(y)dy$$

by Cortes

$$= \mathbb{E}\big(g(Y)\big)$$

- $\boxed{\mathbb{E}\big(X\big) = \mathbb{E}(\mathbb{E}\big(X \mid Y\big))}$

is defined as $g(Y)$

Overview

Continuous random vectors

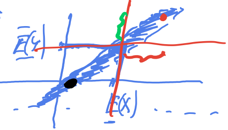
Covariance and correlation

Inequalities

Limit theorems – approximation

Covariance

$\boxed{cov \to 0}$



$\boxed{cov < 0}$

Definition

For r.v.'s X, Y we define their covariance by formula

$$cov(X,Y) = \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big).$$

$= \mathbb{E}\big(\underbrace{XY - \mathbb{E}(X)\cdot Y - X\cdot\mathbb{E}(Y) + \mathbb{E}(X)\cdot\mathbb{E}(Y)}_{\text{sum of r.v.}}\big) = \mathbb{E}(XY) - \mathbb{E}(\mathbb{E}(X)Y) - \mathbb{E}(X\mathbb{E}(Y)) + \mathbb{E}(\underbrace{\mathbb{E}(X)\mathbb{E}(Y)}_{\text{constant}})$

constant

$\mathbb{E}X\cdot\mathbb{E}Y$

Theorem

$$cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

- $var(X) = cov(X,X)$ by def.
- $cov(X, aY + bZ + c) = a\,cov(X,Y) + b\,cov(X,Z)$
- $cov(X,Y) = 0$ if X,Y are independent $\Rightarrow \mathbb{E}(XY) = \mathbb{E}(X)\cdot\mathbb{E}(Y)$
- but not only then

Correlation

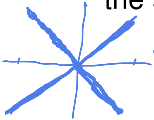
Definition

Correlation of random variables X, Y is defined by

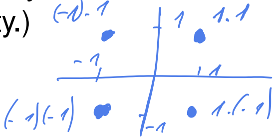
$$\underline{\varrho(X,Y)} = \frac{cov(X,Y)}{\sqrt{\underline{var}(X)\,\underline{var}(Y)}}.$$

- "scaled covariance"
- $-1 \leq \varrho(X,Y) \leq 1$ (exercise)
- Correlation does not imply causation! (In particular, correlation is symmetric.)

  $X = \pm 1$
- OTOH, uncorrelation does not imply independence. (Extreme case: X any r.v., $Y = +X$ or $Y = -X$, both with the same probability.)



$\rho = 0$



$\rho(X,Y) = cov(X,Y) = 0$

indep. r.v.

Variance of a sum $\circledast = \sum_i \sum_j \underbrace{\left(\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\right)}_{cov(X_i, X_j)}$

Theorem

Let $X = \sum_{i=1}^{n} X_i$. *Then*

$$var(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} cov(X_i, X_j) = \sum_{i=1}^{n} var(X_i) + \sum_{i \neq j} cov(X_i, X_j).$$

In particular, if $X_1, \ldots, X_n$ *are independent, then*

$cov(X_i, X_j) = 0$ for $i \neq j$

$$\boxed{var(X) = \sum_{i=1}^{n} var(X_i).}$$

Proof

$var(X) = \mathbb{E}(X \cdot X) - (\mathbb{E}X) \cdot (\mathbb{E}X)$

$= \mathbb{E}\left(\sum_{i=1}^{n} X_i \cdot \sum_{j=1}^{n} X_j\right) - \mathbb{E}\sum X_i \cdot \mathbb{E}\sum X_j$

$\mathbb{E}\left(\sum_i \sum_j X_i X_j\right) - \sum_i \mathbb{E}(X_i) \cdot \sum_j \mathbb{E}(X_j) = \circledast$

Overview

Continuous random vectors

Covariance and correlation

Inequalities

Limit theorems – approximation

Cauchy inequality

Theorem

Let X, Y have finite expectation and variance. Then

$$\boxed{|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}}$$

Proof Cauchy ineq. from L.A.

vectors X, Y .... inner product $\langle X, Y\rangle = \mathbb{E}(X \cdot Y)$

norm $\|X\| = \sqrt{\langle X, X\rangle} = \sqrt{\mathbb{E}(X^2)}$

$|\langle X, Y\rangle| \leq \|X\| \cdot \|Y\|$

- Corollary for correlation: $-1 \leq \varrho(X, Y) \leq 1$

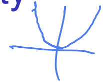
$X' = X - \mathbb{E}X$

$Y' = Y - \mathbb{E}Y$ $\longrightarrow$

$\mathrm{cov}(X, Y) = \mathbb{E}(X'Y') \leq \sqrt{\mathbb{E}(X'^2)\,\mathbb{E}(Y'^2)} = \sqrt{\mathrm{var}(X) \cdot \mathrm{var}(Y)}$

Jensen inequality

Ex. $g(t)=t^2$



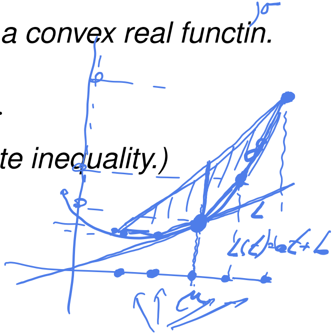
$\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2$

Theorem

Let X have finite expectation and let g be a convex real functin. Then

$$\mathbb{E}\big(g(X)\big) \geq g\big(\mathbb{E}(X)\big).$$

(For concave function we have the opposite inequality.)



Proof $\mu := E(X)$

$g(\mu) = L(\mu)$

$g(t) \geq L(t) \quad \forall t$

$\Downarrow$

$g(X) \geq L(X)$

$\Downarrow$

$\mathbb{E}(g(X)) \geq \mathbb{E}L(X)$; $E(aX+b) = a\mathbb{E}(X)+b$

$= L(\mathbb{E}(X)) = L(\mu) = g(\mu)$

$g(\mathbb{E}X)$

Markov inequality

$B_1 = \{X \geq a\}, \quad B_2 = \{X < a\}$

Theorem

Suppose $X \geq 0$ and $a > 0$. Then

r.v.

average

$$\boxed{P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.}$$

Pf $\mathbb{E}(X) = P(X \geq a) \cdot \underbrace{\mathbb{E}(X \mid X \geq a)}_{\geq a} + \underbrace{P(X<a)}_{\geq 0} \underbrace{\mathbb{E}(X \mid X<a)}_{\geq 0}$

$\geq P(X \geq a) \cdot a$

Ex. We cannot have 51% people older than thr(ee) average. $\geq$ three

$\mu = \mathbb{E}(X)$

$X(\omega) =$ age of person ω

$\Omega =$ set of people, assume uniform prob. on Ω

$a = 2\,\mathbb{E}(X)$

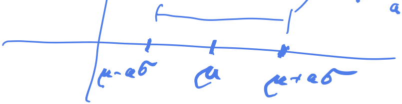
Chebyshev inequality

Theorem

Let X have finite expectation μ and variance σ^2, let $a > 0$. Then

$$\boxed{P(|X-\mu| \geq a\cdot\sigma) \leq \frac{1}{a^2}.}$$

$$P(X \text{ is here}) \geq 1 - \frac{1}{a^2}$$



Markov

$$P(Y \geq a^2\sigma^2) \leq \frac{\mathbb{E}(Y)}{a^2\sigma^2}$$

$$= \frac{\sigma^2}{a^2\sigma^2} = \frac{1}{a^2}$$

Proof $Y = (X-\mu)^2 \geq 0$

Chernoff inequality

Theorem

Let $X = \sum_{i=1}^{n} X_i$*, where* X_i *are i.i.d. attaining* ± 1 *with probability* $1/2$*. Then for* $t > 0$ *we have*

$$P(X \leq -t) = P(X \geq t) \leq e^{-t^2/2\sigma^2},$$

where $\sigma = \sigma_X = \sqrt{n}$*.*

Without proof.

Overview

Strong law of large numbers

Theorem

Let $X_1, \dots, X_n$ be i.i.d. with expectation μ and variance σ^2. Let $S_n = (X_1 + \cdots + X_n)/n$ be the sample mean. Then we have

$$\lim_{n\to\infty} S_n = \mu \quad \text{almost surely (i.e. with probability 1).}$$

We say that sequence S_n converges to μ almost surely.

Monte Carlo integration

How to compute $\int_{x \in A} g(x)dx$?
In particular

$$g(x) = \begin{cases} 1 & \text{for } x \in S \\ 0 & \text{otherwise} \end{cases}$$

... area of a circle

Weak law of large numbers

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. with expectation μ and variance σ^2. Let $S_n = (X_1 + \cdots + X_n)/n$ be the sample mean. Then for every $\varepsilon > 0$ we have

$$\lim_{n \to \infty} P(|S_n - \mu| > \varepsilon) = 0.$$

We say that sequence S_n converges to μ in probability.

Central Limit Theorem

Central Limit Theorem

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. with expectation μ and variance σ^2. Put $Y_n := ((X_1 + \cdots + X_n) - n\mu)/(\sqrt{n} \cdot \sigma)$.
Then $Y_n \xrightarrow{d} N(0,1)$. This means, that if F_n is the cdf of Y_n, then

$$\lim_{n\to\infty} F_n(x) = \Phi(x) \quad \textit{for every } x \in \mathbb{R}.$$

We say that the sequence Y_n converges to $N(0,1)$ in distribution.

Moment generating function

Definition

For a random variable X we let

$$M_X(t) = \mathbb{E}\big(e^{tX}\big).$$

Function $M_X(t)$ is called the moment generating function.

- $M_{Bern(p)}(t) = p \cdot e^t + (1-p)$.
- $M_X(t) = \sum_{n=0}^{\infty} \mathbb{E}\big(X^n\big) \frac{t^n}{n!}$.
- $M_{X+Y}(t) = M_X(t) M_Y(t)$, jsou-li X, Y n.n.v.
- $M_{Bin(n,p)} = (pe^t + 1 - p)^n$
- $M_{N(0,1)} = e^{t^2/2}$
- $M_{Exp(\lambda)} = \frac{1}{1-t/\lambda}$
- If $M_X(t) = M_Y(t)$ on $(-a, a)$ for some $a > 0$, then $X = Y$ a.s.