Statistics Lecture

Claudio Ortelli [1]

[1] Finance Institute
Università della Svizzera italiana

Advanced Learning and Research Institute, 2011

Conditional Distribution and Expectation

Let A and B be two events and $P(B) \neq 0$. The conditional probability of the event A, given that event B is realized, is by definition

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Let X be a random variable and define B to be the event that $X = x$. The conditional probability $P(A \mid X = x)$ of the event A is then

$$P(A \mid X = x) = \frac{P(A \text{ and } X = x)}{P(X = x)} = \frac{P(A \cap [X = x])}{P(X = x)} =$$

provided of course that $P(X = x) \neq 0$.

Conditional Distribution and Expectation

Definition

1) Conditional pmf. Let X and Y be discrete random variables with joint pmf $p(x,y)$. The conditional pmf of Y given X is

$$\begin{aligned} p_{Y|X}(y \mid x) &= P(Y = y \mid X = x) \\ &= \frac{P(Y = y, X = x)}{P(X = x)} = \frac{p(x,y)}{p_x(x)} \end{aligned}$$

if $p_X(x) \neq 0$ and 0 otherwise.

2) The conditional distribution function of a random Variable Y (not necessarily descrete) given a discrete random variable X is

$$F_{Y|X}(y \mid x) = P(Y \leq y \mid X = x) = \frac{P(Y \leq y \text{ and } X = x)}{P(X = x)}$$

for all y and all x such that $P(X = x) \neq 0$.

Conditional Distribution and Expectation

Example

- Server cluster with two servers labeled A and B.
- Incoming jobs are independently routed to A and B with probability p and $q = 1 - p$, respectively.
- The number X of arriving jobs per unit of time is Poisson distributed with intensity λ.
- Determine the number of jobs, Y, received by server A, per unit of time.

$$P_{Y|X}(k,n) = \begin{cases} P_{Y|X}(Y=k, X=n) = \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}, & 0 \leq k \leq n. \\ 0 & \text{otherwise.} \end{cases}$$

Conditional Distribution and Expectation

Example

(Continued) Recall that $P(X=n)=e^{-\lambda}\lambda^n/n!$ so that

$$
\begin{aligned}
p_Y(k) &= \sum_{n=k}^{\infty} p_{Y|X}(k \mid n)p_X(n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \lambda^k p^k e^{-\lambda} \sum_{n=k}^{\infty} \binom{n}{k} \frac{1}{n!} q^{n-k}\lambda^{n-k} \\
&= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{n=k}^{\infty} \frac{(q\lambda)^{n-k}}{(n-k)!}
\end{aligned}
$$

so that finally $p_Y(k)=\frac{(\lambda p)^k}{k!}e^{-\lambda}e^{q\lambda}=\frac{(\lambda p)^k}{k!}e^{-\lambda p}$, i.e. Y is Poisson distributed with intensity λp.

Conditional Distribution and Expectation

If X is a continuous random variable then $P(X=x)=0$ for all $x \in \mathbb{R}$ so that the previous definition $\frac{P(Y=y,X=x)}{P(X=x)}$ of conditional probability is not satisfactory.
However when X and Y are jointly continuous we can define the conditional pdf of Y given X:

Definition

Let X and Y be continuous r.v. with joint pdf $f(x,y)$. The conditional density $f_{Y|X}$ is

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{f(x,y)}{f_X(x)}, & \text{if } 0 < f_X(x) < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Conditional Distribution and Expectation

From the definition of conditional density it follows that

$$f(x,y) = f_{Y|X}(y \mid x) f_X(x) = f_{X|Y}(x \mid y) f_Y(y),$$

and if X and Y are independent, then

$$f(x,y) = f_X(x) f_Y(y).$$

Furthermore,

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

which is the continuous analog of the thorem of total probability.

Conditional Distribution and Expectation

- The conditional pdf can be used to obtain the conditional probability:

$$P(a \leq Y \leq b \mid X = x) = \int_a^b f_{Y|X}(y \mid x)dy, \qquad a \leq b.$$

- The conditional distribution function is defined analogously

$$\begin{aligned} F_{Y|X}(y \mid x) &= P(Y \leq y \mid X = x) \\ &= \frac{\int_{-\infty}^{y} f(x,t)dt}{f_X(x)} \\ &= \int_{-\infty}^{y} f_{Y|X}(t \mid x)dt. \end{aligned}$$

Example

Consider a series system of two *independent* components with respective lifetime distributions $X \sim EXP(\lambda_1)$ and $Y \sim EXP(\lambda_2)$. We are interested in the probability of envent A that component 2 causes the system failure, i.e.

$$P(A) = P(X \geq Y).$$

The conditional pdf is $F_{X|Y}(t,t) = P(X \leq t \mid Y = t) = F_X(t)$ by the independence of X and Y. By the total prob. theorem (continuous version)

$$\begin{aligned} P(A) &= \int_0^\infty P(X \geq t, Y = t) f_Y(t) dt \\ &= \int_0^\infty [1 - F_X(t)] f_Y(t) dt = \frac{\lambda_2}{\lambda_1 + \lambda_2}. \end{aligned}$$

Conditional Distribution and Expectation

Exercise

Consider the three-dimensional vector $X = (X_1, X_2, X_3)$ *having the following joint density function*

$$f_X(x_1, x_2, x_3) = \begin{cases} 6x_1 x_2^2 x_3, & \text{if } 0 \leq x_1 \leq 1, 0 \leq x_x \leq 1, 0 \leq x_3 \leq \sqrt{2}. \\ 0, & \text{otherwise.} \end{cases}$$

1. *Compute the conditional density functions* $f_{X_1, X_2 | X_3}(x_1, x_2 \mid x_3)$ *and* $f_{X_3 | X_1}(x_3 \mid x_1)$.
2. *Verify if the three random variables* X_1, X_2, X_3 *are independent.*

Exercise

X_1 *and* X_2 *are independent r. v. with Poisson distribution, having respective parameters* α_1 *and* α_2. *Show that the conditional* pmf *of* X_1 *given* $X_1 + X_2$, $p_{X_1 | X_1 + X_2}(X_1 = x_1 \mid X_1 + X_2 = y)$, *is binomial. Determine its parameters.*

Exercise

Let the execution times X and Y of two independent parallel processes be uniformly distributed over $(0, t_X)$ and $(0, t_Y)$, respectively, with $t_X \leq t_Y$. Find the probability that the former process finishes execution before the later.

- Consider a file server whose workload may be divided into r distinct classes.
- For a job of class i ($1 \leq i \leq r$) the CPU time is exponentially distributed with parameter λ_i.
- Let Y denote the service time of a job and let X be the job class. Then

$$f_{Y|X}(y \mid i) = \lambda_i e^{-\lambda_i y}, \qquad y > 0.$$

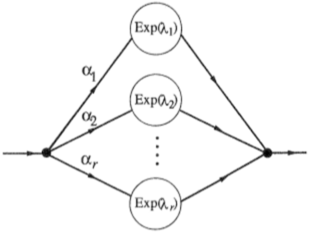
- Assume that the probability $p_X(i)$ that a randomly chosen job belongs to class i is equal to $\alpha_i > 0$. It follows $\sum_{i=1}^{r} \alpha_i = 1$.

The joint density of X and Y is then

$$f(i,y) = f_{Y|X}(y \mid i) p_X(i) = \alpha_i \lambda_i e^{-\lambda_i y}, \qquad y > 0.$$

The marginal density of Y is then

$$f_Y(y) = \sum_{i=1}^{r} f(i,y) = \sum_{i=1}^{r} \alpha_i \lambda_i e^{-\lambda_i y}, \qquad y > 0, \text{ i.e.}$$



Y has an $r-$stage hyperexponential distribution!

In general the conditional distribution of Y does not have to be exponential! Denoting $f_{Y|X}(y \mid i) = f_{Y_i}(y)$ and $F_{Y|X}(y \mid i) = F_i(y)$ then the unconditional pdf of Y is

$$f_Y(y) = \sum_{i=1}^{r} \alpha_i f_i(y)$$

and the unconditional CDF of Y is

$$F_Y(y) = \sum_{i=1}^{r} \alpha_i F_i(y).$$

Applying the definition of the mean and higher moments we obtain

$$\begin{aligned} E[Y] &= \sum_{i=1}^{r} \alpha_i E[Y_i], \\ E[Y^k] &= \sum_{i=1}^{r} \alpha_i E[Y_i^k]. \end{aligned}$$

Conditional Distribution and Expectation

- If X and Y are continuous random variables, we can for instance compute the conditional density $f_{Y|X}$.
- Since $f_{Y|X}$ has all properties of a density function of a continuous random variable, we can talk about its moments.
- Its mean (if exists) is called the conditional expectation of Y given $X = x$ and is denoted $E[Y \mid X = x]$ or $E[Y \mid x]$:

$$E[Y \mid x] \quad = \quad \begin{cases} \int_{-\infty}^{\infty} y f(y \mid x) dy, & \text{if } 0 < f(x) < \infty \\ 0 & \text{otherwise.} \end{cases}$$

- In case the random variables X and Y are discrete, $E[Y \mid x]$ is defined as

$$E[Y \mid X = x] = \sum_y y P(Y = y \mid X = x) = \sum_y y p_{Y|X}(y \mid x).$$

Similar arguments hold when X and Y are discrete. The conditional expectation is then defined as

$$E[Y \mid X = x] = \sum_y yP(Y = y \mid X = x) = \sum_y y p_{Y|X}(y \mid x).$$

Definition

The quantity

$$m(x) = E[Y \mid x]$$

considered as a function of x is known as the *regression function* of Y on X.

Conditional Distribution and Expectation

Definition

The conditional expectation of a function $\phi(Y)$ is defined as

$$E[\phi(Y) \mid X = x] = \begin{cases} \int_{-\infty}^{\infty} \phi(y) f_{Y|X}(y \mid x) dy, & \text{if } Y \text{ is continuous,} \\ \sum_i \phi(y_i) p_{Y|X}(y_i \mid x), & \text{if } Y \text{ is discrete.} \end{cases}$$

We may take expectation of the regression function to obtain the unconditional expectation of $\phi(Y)$

$$E[\phi(Y)] = \begin{cases} \sum_x E[\phi(Y) \mid X = x] p_X(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} E[\phi(Y) \mid X = x] f_X(x) dx, & \text{if } X \text{ is continuous.} \end{cases}$$

This last formula is known as the **theorem of total expectation**.

Limit Theorems

Theorem

(Chebyshev) Let X be a random variable with expected value μ and finite variance $\sigma^2 < \infty$. The, for all $t > 0$, the following inequality holds

$$P(\mid X - \mu \mid \geq t) \leq \frac{\sigma^2}{t^2}.$$

Definition

(Convergence in probability) Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables. We say that the sequence converges in probability to $c \in \mathbb{R}$, write $X_n \underset{p}{\rightarrow} c$ or $p\lim X_n = c$, if, for all $\varepsilon > 0$

$$\lim_{n \to \infty} P(\mid X_n - c \mid \geq \varepsilon) = 0.$$

Limit Theorems

Theorem

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables with common expectation μ and finite variance $\sigma_n^2 < \infty$. If $\lim_{n\to\infty} \sigma_n^2 = 0$, then

$$X_n \underset{p}{\to} \mu.$$

Proof.

Apply Chebyshev inequality. □

Example

Let $\{X_n\}_{n\in\mathbb{N}}$ be an $i.i.d. \sim (\mu, \sigma^2)$ (independent and identically distributed) sequence of random variables. Define the sequence $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. From the linearity of the expectation and the properties of the variance we know that $\bar{X}_n \sim (\mu, \frac{\sigma^2}{n})$. The sequence $\{\bar{X}_n\}_{n\in\mathbb{N}}$ converges in probability to μ: $p\lim \bar{X}_n = \mu$.

Limit Theorems

The result in the previous Example is also known as the **weak law of large numbers** (WLLN). In order for the WLLN to apply the existence of the second moment (the variance) is not required. The WLLN holds just under the assumption that the $\{X_n\}_{n\in\mathbb{N}}$ $i.i.d.$ sequence have finite expected value μ.

Theorem

Consider two sequences $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ of random variables converging in probability to $a<\infty$ and $b<\infty$, respectively. Then

1. $p\lim(X_n+Y_n)=p\lim X_n+p\lim Y_n=a+b.$
2. $p\lim(X_n\cdot Y_n)=p\lim X_n\cdot p\lim Y_n=a\cdot b.$
3. $b\neq 0,$
$$p\lim\left(\frac{X_n}{Y_n}\right)=\frac{p\lim X_n}{p\lim Y_n}=\frac{a}{b}.$$
4. Function g continuous in a : $p\lim g(X_n)=g(p\lim X_n)=g(a).$

Definition

(Standardization) Let X be a random variable with expected value μ and finite variance σ^2. The location and scale trasform

$$Z = \frac{X - \mu}{\sigma}$$

defines the standardization of X. From the properties of expectation it is straightforward to prove that $Z \sim (0, 1)$.

Example

Let $\{X_n\}_{n \in \mathbb{N}}$ be an indipendent sequence of random variables with $X_n \sim (\mu, \sigma^2)$ for all $n \in \mathbb{N}$. Define the sequence $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$Z_n = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \sim (0, 1) \text{ for all } n \in \mathbb{N}.$$

Limit Theorems

Theorem

The Central Limit Theorem (CLT). Let $\{X_n\}_{n\in\mathbb{N}}$ *be independent random variables with a finite mean* $E[X_n] = \mu_n$ *and a finite variance* $Var(X_n) = \sigma_n^2$. *Define the normalized random variable*

$$Z_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

so that $E[Z_n] = 0$ *and* $Var(Z_n) = 1$ *for all* n. *Then under regularity conditions the limiting distribution of* Z_n *is standard normal, denoted* $Z_n \to N(0,1)$, *i.e.*

$$\lim_{n\to\infty} F_{Z_n}(t) = \lim_{n\to\infty} P(Z_n \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Remark: the special condition X_n *independent with* $Var(X_n) = \sigma^2$ *for all* n *is sufficient for the CTL to apply.*

Limit Theorems

exercises ...

Parameter estimation

The object under study is

1. the probability distribution function F of a random experiment or random variable X, or
2. the statistical distribution function F of a given population of individuals, users, devices, ...

We assume that F is known up to a vector of unknown parameters θ.

Definition

The family of distributions $\mathscr{P} = \{F_\theta\}_{\Theta \subseteq \mathbb{R}^n}$, $n \in \mathbb{N}$ finite, is called parametric model. The parametric model is usually specified in terms of probability mass or density functions.

Parameter estimation

Example

The Poisson family of distributions is parametrized by a single parameter $\lambda > 0$

$$\mathscr{P} = \left\{ p_\lambda(j) = \frac{\lambda^j}{j!} \exp^{-\lambda}, j = 0, 1, \ldots \mid \lambda > 0 \right\}.$$

The Normal family is parametrized by two parameters $\theta = (\mu, \sigma)$

$$\mathscr{P} = \left\{ f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{1}{2\sigma^2}(x-\mu)^2} \mid \mu \in \mathbb{R}, \sigma > 0 \right\}.$$

The Logistic distribution is defined by the following distribution function where $\theta = (\mu, \sigma)$:

$$\mathscr{P} = \left\{ F_\theta(x) = \frac{1}{1 + \exp^{-(x-\mu)/\beta}} \mid \mu \in \mathbb{R}, \sigma > 0 \right\}.$$

Point estimation problem

Estimator

Estimate

Unbiased estimator