Lecture 2

Last updated by Serik Sagitov: November 23, 2012

Abstract

A course based on the book Probabilities and Random Processes by Geoffrey Grimmett and David Stirzaker. Chapter 3. Discrete random variables. Chapter 4. Continuous random variables.

1 Expectation

The expected value of X is

$$\mathbb{E}(X) = \int_\Omega X(\omega)\mathbb{P}(d\omega).$$

A discrete r.v. X with a finite number of possible values is a simple r.v. in that

$$X = \sum_{i=1}^n x_i I_{A_i}$$

for some partition $A_1, \ldots, A_n$ of Ω. In this case the meaning of the expectation is obvious

$$\mathbb{E}(X) = \sum_{i=1}^n x_i \mathbb{P}(A_i).$$

For any non-negative r.v. X there are simple r.v. such that $X_n(\omega) \nearrow X(\omega)$ for all $\omega \in \Omega$, and the expectation is defined as a possibly infinite limit $\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n)$.

Any r.v. X can be written as a difference of two non-negative r.v. $X^+ = X \vee 0$ and $X^- = -X \wedge 0$. If at least one of $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ is finite, then $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$, otherwise $\mathbb{E}(X)$ does not exist.

Example 1 *A discrete r.v. with the probability mass function* $f(k) = \frac{1}{2k(k-1)}$ *for* $k = -1, \pm 2, \pm 3, \ldots$ *has no expectation.*

For a discrete r.v. X with mass function f and any function g

$$\mathbb{E}(g(X)) = \sum_x g(x)f(x).$$

For a continuous r.v. X with density f and any measurable function g

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

In general

$$\mathbb{E}(X) = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_{-\infty}^{\infty} x\mathbb{P}_X(dx) = \int_{-\infty}^{\infty} xdF(x) = \int_{-\infty}^{\infty} (1 - F(x))dx.$$

Example 2 *Turn to the example in Lecture 1 of a random variable X with the Cantor distribution. A sequence of simple r.v. monotonely converging to X*

$$\begin{aligned}
&X_1(\omega) = 0, \quad \mathbb{E}(X_0) = 0,\\
&X_2(\omega) = (1/2)I_{\{[1/3,1]\}}(\omega), \quad \mathbb{E}(X_1) = 1/4,\\
&X_3(\omega) = (1/4)I_{\{[1/9,1/3]\}}(\omega) + (1/2)I_{\{[1/3,4/9]\}}(\omega) + (3/4)I_{\{[4/9,1]\}}(\omega), \quad \mathbb{E}(X_2) = 3/8, \ldots
\end{aligned}$$

gives $\mathbb{E}(X) = 1/2$.

Cauchy-Schwartz inequality: for r.v. X and Y

$$\big(\mathbb{E}(XY)\big)^2 \le \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

with equality if only if $aX + bY = 1$ a.s. for some non-trivial pair of constants (a, b). Variance, standard deviation, covariance and correlation

$$\begin{aligned} \mathrm{var}(X) &= \mathbb{E}\big(X - \mathbb{E}X\big)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2, \quad \sigma_X = \sqrt{\mathrm{var}(X)}, \\ \mathrm{cov}(X,Y) &= \mathbb{E}\big(X - \mathbb{E}X\big)\big(Y - \mathbb{E}Y\big) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y), \\ \rho(X,Y) &= \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}. \end{aligned}$$

Definition 3 *Random variables $(X_1, \ldots, X_n)$ are called independent if for any $(x_1, \ldots, x_n)$*

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \ldots \mathbb{P}(X_n \le x_n).$$

In the jointly continuous case this equivalent to

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n).$$

2 Conditional expectation

Definition 4 *For a pair of discrete random variables (X, Y) the conditional expectation $\mathbb{E}(Y|X)$ is defined as $\psi(X)$, where*

$$\psi(x) = \sum_y y\mathbb{P}(Y = y|X = x).$$

Definition 5 *Consider a pair of random variables (X, Y) with joint density $f(x, y)$, marginal densities*

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

and conditional densities

$$f_1(x|y) = \frac{f(x,y)}{f_2(y)}, \qquad f_2(y|x) = \frac{f(x,y)}{f_1(x)}$$

The conditional expectation $\mathbb{E}(Y|X)$ is defined as $\psi(X)$, where

$$\psi(x) = \int_{-\infty}^{\infty} y f_2(y|x).$$

Properties of conditional expectations

- linearity: $\mathbb{E}(aY + bZ|X) = a\mathbb{E}(aY|X) + b\mathbb{E}(Z|X)$ for any constants $(a, b$,
- pull-through property: $\mathbb{E}(Yg(X)|X) = g(X)\mathbb{E}(Y|X)$ for any measurable function $g(x)$,
- tower property: $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$ or more generally $\mathbb{E}(\mathbb{E}(Y|X,Z)|X) = \mathbb{E}(Y|X)$.

3 Multinomial distribution

De Moivre trials: each trial has r possible outcomes with probabilities $(p_1, \ldots, p_r)$. Consider n such independent trials and let $(X_1, \ldots, X_r)$ be the counts of different outcomes. Multinomial distribution $\mathrm{Mn}(n, p_1, \ldots, p_r)$

$$\mathbb{P}(X_1 = k_1, \ldots, X_r = k_r) = \frac{n!}{k_1! \ldots k_r!} p_1^{k_1} \ldots p_r^{k_r}.$$

Marginal distributions $X_i \sim \mathrm{Bin}(n, p_i)$, also

$$(X_1 + X_2, X_3 \ldots, X_r) \sim \mathrm{Mn}(n, p_1 + p_2, p_3, \ldots, p_r).$$

Conditionally on X_1

$$(X_2, \ldots, X_r) \sim \mathrm{Mn}(n - X_1, \frac{p_2}{1-p_1}, \ldots, \frac{p_r}{1-p_1}),$$

so that $(X_i|X_j) \sim \mathrm{Bin}(n - X_j, \frac{p_i}{1-p_j})$ and $\mathbb{E}(X_i|X_j) = (n - X_j)\frac{p_i}{1-p_j}$. It follows

$$\begin{aligned}\mathbb{E}(X_iX_j) &= \mathbb{E}(\mathbb{E}(X_iX_j|X_j)) \\ &= \mathbb{E}(X_j\mathbb{E}(X_i|X_j)) = \mathbb{E}(nX_j - X_j^2)\frac{p_i}{1-p_j} \\ &= (n^2p_j - np_j(1-p_j) + n^2p_j^2)\frac{p_i}{1-p_j} = n(n-1)p_ip_j\end{aligned}$$

and $\mathrm{cov}(X_i, X_j) = -np_ip_j$ so that

$$\rho(X_i, X_j) = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}.$$

4 Multivariate normal distribution

Bivariate normal distribution with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{(\frac{x-\mu_1}{\sigma_1})^2 - 2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2}) + (\frac{y-\mu_2}{\sigma_2})^2}{2(1-\rho^2)}\right\}.$$

Marginal distributions

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \quad f_2(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},$$

and conditional distributions

$$\begin{aligned}f_1(x|y) = \frac{f(x,y)}{f_2(y)} &= \frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{(x-\mu_1-\frac{\rho\sigma_1}{\sigma_2}(y-\mu_2))^2}{2\sigma_1^2(1-\rho^2)}\right\}, \\ f_2(y|x) = \frac{f(x,y)}{f_1(x)} &= \frac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{(y-\mu_2-\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1))^2}{2\sigma_2^2(1-\rho^2)}\right\}.\end{aligned}$$

The covariance matrix of a random vector $(X_1, \ldots, X_n)$ with means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$

$$\mathbf{V} = \mathbb{E}\big(\mathbf{X} - \boldsymbol{\mu}\big)^{\mathrm{t}}\big(\mathbf{X} - \boldsymbol{\mu}\big) = \|\mathrm{cov}(X_i, X_j)\|$$

is symmetric and nonnegative-definite. For any vector $\mathbf{a} = (a_1, \ldots, a_n)$ the r.v. $a_1X_1 + \ldots + a_nX_n$ has mean $\mathbf{a}\boldsymbol{\mu}^{\mathrm{t}}$ and variance

$$\mathrm{var}(a_1X_1 + \ldots + a_nX_n) = \mathbb{E}\big(\mathbf{a}\mathbf{X}^{\mathrm{t}} - \mathbf{a}\boldsymbol{\mu}^{\mathrm{t}}\big)\big(\mathbf{X}\mathbf{a}^{\mathrm{t}} - \boldsymbol{\mu}\mathbf{a}^{\mathrm{t}}\big) = \mathbf{a}\mathbf{V}\mathbf{a}^{\mathrm{t}}.$$

A multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and covariance matrix $\mathbf{V}$ has density

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det\mathbf{V}}} e^{-(\mathbf{x}-\boldsymbol{\mu})\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{t}}}.$$

For any vector $(a_1, \ldots, a_n)$ the r.v. $a_1X_1 + \ldots + a_nX_n$ is normally distributed. Application in statistics: in the IID case: $\boldsymbol{\mu} = (\mu, \ldots, \mu)$ and $\mathbf{V} = \mathrm{diag}\{\sigma^2, \ldots, \sigma^2\}$ the sample mean and sample variance

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n}, \quad s^2 = \frac{(X_1 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n-1}$$

are independent and $\frac{\sqrt{n}(\bar{X}-\mu)}{s}$ has a t-distribution with $n-1$ degrees of freedom.

If Y and Z are independent r.v. with standard normal distribution, their ratio $X = Y/Z$ has a Cauchy distribution with density

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

In the Cauchy distribution case the mean is undefined and $\bar{X} \stackrel{d}{=} X$. Cauchy and normal distributions are examples of stable distributions. The Cauchy distribution provides with a counterexample for the law of large numbers.

5 Sampling from a distribution

Computers generate pseudo-random numbers $U_1, U_2, \ldots$ which we consider as IID r.v. with $\mathrm{U}_{[0,1]}$ distribution.

Inverse transform sampling: if F is a cdf and $U \sim \mathrm{U}_{[0,1]}$, then $X = F_{-1}(U)$ has cdf F. It follows from

$$\{F_{-1}(U) \le x\} = \{U \le F(x)\}.$$

Examples

- Bernoulli distribution $X = I_{\{U \le p\}}$,
- binomial sampling: $S_n = X_1 + \ldots + X_n$, $X_k = I_{\{U_k \le p\}}$,
- exponential distribution $X = -\ln(U)/\lambda$,
- gamma sampling: $S_n = X_1 + \ldots + X_n$, $X_k = -\ln(U_k)/\lambda$.

Rejection sampling. Suppose that we know how to sample from density $g(x)$ but we want to sample from density $f(x)$ such that $f(x) \le ag(x)$ for some $a > 0$. Algorithm

- sample x from $g(x)$ and u from $\mathrm{U}_{[0,1]}$,
- if $u \le \frac{f(x)}{ag(x)}$, accept x as a realization of sampling from $f(x)$,
- if not, reject the value of x and repeat the sampling step.

Proof. Let Z and U be independent, Z has density $g(x)$ and $U \sim \mathrm{U}_{[0,1]}$. Then

$$\mathbb{P}\Big(Z \le x \Big| U \le \frac{f(Z)}{ag(Z)}\Big) = \frac{\int_{-\infty}^{x} \mathbb{P}\Big(U \le \frac{f(y)}{ag(y)}\Big) g(y)dy}{\int_{-\infty}^{\infty} \mathbb{P}\Big(U \le \frac{f(y)}{ag(y)}\Big) g(y)dy} = \int_{-\infty}^{x} f(y)dy.$$

6 Simple random walks

Let $S_n = a + X_1 + \ldots + X_n$ where $X_1, X_2, \ldots$ are IID r.v. taking values 1 and -1 with probabilities p and $q = 1-p$. This Markov chain is homogeneous both in space and time. We have $S_n = 2Z_n - n$, with $Z_n \sim \mathrm{Bin}(n,p)$. Symmetric random walk if $p = 0.5$. Drift upwards $p > 0.5$ or downwards $p < 0.5$ like in casino.

The ruin probability $p_k = p_k(N)$: your starting capital k against casino's $N - k$. The difference equation

$$p_k = p \cdot p_{k+1} + q \cdot p_{k-1}, \quad p_N = 0, \quad p_0 = 1$$

gives

$$p_k(N) = \begin{cases} \frac{(q/p)^N - (q/p)^k}{(q/p)^N - 1}, & \text{if } p \ne 0.5, \\ \frac{N-k}{N}, & \text{if } p = 0.5. \end{cases}$$

Start from zero and let τ_b be the first hitting time of b, then for $b > 0$

$$\mathbb{P}(\tau_{-b} < \infty) = \lim_{N\to\infty} p_b(N) = \begin{cases} 1, & \text{if } p \le 0.5, \\ (q/p)^b, & \text{if } p > 0.5, \end{cases}$$

and

$$\mathbb{P}(\tau_b < \infty) = \begin{cases} 1, & \text{if } p \ge 0.5, \\ (p/q)^b, & \text{if } p < 0.5. \end{cases}$$

The mean number $D_k = D_k(N)$ of steps before hitting either 0 or N. The difference equation

$$D_k = p \cdot (1 + D_{k+1}) + q \cdot (1 + D_{k-1}), \quad D_0 = D_N = 0$$

gives

$$D_k(N) = \begin{cases} \frac{1}{q-p}\left[k - N \cdot \frac{1-(q/p)^k}{1-(q/p)^N}\right], & \text{if } p \ne 0.5, \\ k(N-k), & \text{if } p = 0.5. \end{cases}$$

If $p < 0.5$, then the expected ruin time is computed as $D_k(N) \to \frac{k}{q-p}$ as $N \to \infty$.

There are

$$N_n(a,b) = \binom{n}{k_a}, \quad k_a = \frac{n+b-a}{2}$$

paths from a to b in n steps. Each path has probability $p^k q^{n-k}$, $k = k_a$. Thus

$$\mathbb{P}(S_n = b|S_0 = a) = \binom{n}{k} p^k q^{n-k}, \quad k = k_a.$$

In particular, $\mathbb{P}(S_{2n} = a|S_0 = a) = \binom{2n}{n}(pq)^n$. Reflection principle: the number of n-paths visiting r is

$$N_n^r(a,b) = N_n(2r-a,b), \quad a \ge r, b \ge r,$$
$$N_n^r(a,b) = N_n(a,2r-b), \quad a < r, b < r.$$

Ballot theorem: if $b > 0$, then the number of n-paths $0 \to b$ not revisiting zero is

$$\begin{aligned} N_{n-1}(1,b) - N_{n-1}^0(1,b) &= N_{n-1}(1,b) - N_{n-1}(-1,b) \\ &= \binom{n-1}{k_0-1} - \binom{n-1}{k_0} = (b/n)N_n(0,b). \end{aligned}$$

Thus (by default we will assume $S_0 = 0$)

$$\begin{aligned} &\mathbb{P}(S_1 > 0, \dots S_{n-1} > 0|S_n = b) = \frac{b}{n}, \quad b > 0, \\ &\mathbb{P}(S_1 \ne 0, \dots S_{n-1} \ne 0, S_n = b) = \frac{|b|}{n}\mathbb{P}(S_n = b), \\ &\mathbb{P}(S_1 \ne 0, \dots S_n \ne 0) = n^{-1}\mathbb{E}|S_n|. \end{aligned}$$

It follows that for $p = 0.5$ we have (see 3.10.22)

$$\mathbb{P}(S_1 \ne 0, \dots S_{2m} \ne 0) = \mathbb{P}(S_{2m} = 0). \tag{1}$$

For the maximum $M_n = \max\{S_0, \dots, S_n\}$ using $N_n^r(0,b) = N_n(0,2r-b)$ for $r > b$ and $r > 0$ we get

$$\mathbb{P}(M_n \ge r, S_n = b) = (q/p)^{r-b}\mathbb{P}(S_n = 2r-b),$$

implying for $b > 0$

$$\mathbb{P}(S_1 < b, \dots S_{n-1} < b, S_n = b) = \frac{b}{n}\mathbb{P}(S_n = b).$$

The obtained equality

$$\mathbb{P}(S_1 > 0, \dots S_{n-1} > 0, S_n = b) = \mathbb{P}(S_1 < b, \dots S_{n-1} < b, S_n = b)$$

can be explained in terms of the reversed walk also starting at zero: the initial walk comes to b without revisiting zero means that the reversed walk reaches its maximum on the final step.

The first hitting time τ_b has distribution

$$\mathbb{P}(\tau_b = n) = \frac{|b|}{n}\mathbb{P}(S_n = b), \quad n > 0.$$

The mean number of visits of $b \neq 0$ before revisiting zero

$$\mathbb{E}\sum_{n=1}^{\infty} I_{\{S_1\neq 0,\dots S_{n-1}\neq 0, S_n=b\}} = \sum_{n=1}^{\infty}\mathbb{P}(\tau_b = n) = \mathbb{P}(\tau_b < \infty).$$

Arcsine law for the last visit to the origin. Let $p = 0.5$, $S_0 = 0$, and T_{2n} be the time of the last visit to zero up to time $2n$. Then

$$\mathbb{P}(T_{2n} \le 2xn) \to \int_0^x \frac{dy}{\pi\sqrt{y(1-y)}} = \frac{2}{\pi}\arcsin\sqrt{x}, \quad n \to \infty.$$

Proof*. Using (1) we get

$$\begin{aligned}\mathbb{P}(T_{2n} = 2k) &= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2k+1} \neq 0, \dots, S_{2n} \neq 0 | S_{2k} = 0)\\ &= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2(n-k)} = 0),\end{aligned}$$

and it remains to apply Stirling's formula.

Arcsine law for sojourn times. Let $p = 0.5$, $S_0 = 0$, and T_{2n}^+ be the number of time intervals spent on the positive side up to time $2n$. Then $T_{2n}^+ \stackrel{d}{=} T_{2n}$.
Proof*. First using

$$\mathbb{P}(S_1 > 0, \dots, S_{2n} > 0) = \mathbb{P}(S_1 = 1, S_2 \ge 1, \dots, S_{2n} \ge 1) = \frac{1}{2}\mathbb{P}(T_{2n}^+ = 2n)$$

and (1) observe that

$$\mathbb{P}(T_{2n}^+ = 0) = \mathbb{P}(T_{2n}^+ = 2n) = \mathbb{P}(S_{2n} = 0).$$

Then by induction over n one can show that

$$\mathbb{P}(T_{2n}^+ = 2k) = \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2(n-k)} = 0)$$

for $k = 1, \dots, n-1$, applying the following useful relation

$$\mathbb{P}(S_{2n} = 0) = \sum_{k=1}^{n}\mathbb{P}(S_{2(n-k)} = 0)\mathbb{P}(\tau_0 = 2k),$$

where τ_0 is the time of first return to zero.