3. Expectations

1. Expectation

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We define the *expectation* $\mathbf{E}(X)$ of X by

$$\mathbf{E}(X) = \int X \, d\mathbf{P}$$

More generally, define for $f(X) = f \circ X$ with a measurable $f : \mathbf{R} \to \mathbf{R}$

$$\mathbf{E}f(X) = \int f(X) \, d\mathbf{P}$$

to be the expectation of a random variable $f(X)$. It is straightforward to show that $f(X)$ is indeed a random variable. Note for any $A \in \mathcal{B}(\mathbf{R})$ that

$$(f \circ X)^{-1}(A) = X^{-1}(f^{-1}(A)) \in \mathcal{F}$$

since f is assumed to be measurable and $f^{-1}(A) \in \mathcal{B}(\mathbf{R})$.

The following theorem tells us how to compute expectations.

Theorem 1

$$\int f(X) \, d\mathbf{P} = \int f \, d\mathrm{P}_X = \int f p_X \, d\mu$$

Proof First we will show that the stated result holds for simple functions. Therefore, we set

$$f = \sum_{k=1}^{n} c_k \mathrm{I}(A_k)$$

so that

$$f(X) = \sum_{k=1}^{n} c_k \mathrm{I}(X^{-1}(A_k))$$

It follows that

$$\begin{aligned} \int f(X) \, d\mathbf{P} &= \sum_{k=1}^{n} c_k \mathbf{P}(X^{-1}(A_k)) \\ &= \sum_{k=1}^{n} c_k \mathrm{P}_X(A_k) \\ &= \int f \, d\mathrm{P}_X \end{aligned}$$

Moreover, we have

$$\begin{aligned}\int f\,d\mathrm{P}_X &= \sum_{k=1}^{n} c_k \mathrm{P}_X(A_k) \\ &= \sum_{k=1}^{n} c_k \left(\int_{A_k} p_X d\mu \right) \\ &= \int f p_X d\mu\end{aligned}$$

Note that

$$\int_{A_k} p_X d\mu = \int \mathrm{I}(A_k) p_X d\mu$$

and that an integral can be interchanged with a finite summation.

Thus we have shown that the stated result holds for simple functions. For a general nonnegative function f, we can choose a sequence of simple functions $\{f_n\}$ such that $f_n \uparrow f$. It then follows from the monotone convergence theorem that the stated result applies to f. Finally, for any measurable function f, we write $f = f^+ - f^-$ to show that the stated result holds. ■

Remark All the properties of the integral, including linearity, apply to the expectation operation.

Expectations of some special functions have special names and meaning. In particular, we call

$$\mu = \mathbf{E}(X) \quad \text{and} \quad \sigma^2 = \mathrm{var}(X) = \mathbf{E}(X - \mu)^2$$

the *mean* and the *variance*, respectively, of the random variable X. Moreover, if we let X and Y be two random variables with means μ_x and μ_y, then

$$\sigma_{xy} = \mathrm{cov}(X, Y) = \mathbf{E}(X - \mu_x)(Y - \mu_y)$$

and

$$\rho_{xy} = \mathrm{cor}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$$

are called, respectively, *covariance* and *correlation* of the random variables X and Y.

Remarks The following are their properties.

(a) $\text{var}(aX) = a^2\text{var}(X)$ and $\text{var}(X + b) = \text{var}(X)$.

(b) $\text{cov}(a_1X_1 + a_2X_2, Y) = a_1\text{cov}(X_1, Y) + a_2\text{cov}(X_2, Y)$. Notice also that

$$\text{cov}(X, b_1Y_1 + b_2Y_2) = b_1\text{cov}(X, Y_1) + b_2\text{cov}(X, Y_2)$$

due to symmetry $\text{cov}(X, Y) = \text{cov}(Y, X)$.

2. Expectational Inequalties

In this section, we derive some useful expectational inequalities. Again, we let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

Theorem 2 (Chebyshev) *Let* $\varepsilon > 0$. *Then*

$$\mathbf{P}\{|X| \geq \varepsilon\} \leq \frac{\mathbf{E}|X|^k}{\varepsilon^k}$$

Proof Notice that

$$\varepsilon^k \text{I}\{|X| \geq \varepsilon\} \leq |X|^k$$

Take expectation on both sides to get the stated result. ∎

Remark We have as a special case of Chebyshev's inequality

$$\mathbf{P}\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

where μ and σ^2 are the mean and variance of X. The inequality thus yields an upper bound for tail probabilities of a random variable with finite variance.

Let X and Y be random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then we have

Theorem 3 (Cauchy-Schwartz)

$$(\mathbf{E}XY)^2 \leq (\mathbf{E}X^2)(\mathbf{E}Y^2)$$

Proof Write

$$Y = \frac{\mathbf{E}XY}{\mathbf{E}X^2}X + \left(Y - \frac{\mathbf{E}XY}{\mathbf{E}X^2}X\right)$$

Squaring and taking expectations on both sides yields

$$\mathbf{E}Y^2 = \mathbf{E}\left(\frac{\mathbf{E}XY}{\mathbf{E}X^2}X\right)^2 + \mathbf{E}\left(Y - \frac{\mathbf{E}XY}{\mathbf{E}X^2}X\right)^2$$

Notice that the expectation of the cross product term vanishes. We have that

$$\mathbf{E}Y^2 \geq \frac{(\mathbf{E}XY)^2}{\mathbf{E}X^2}$$

from which the stated result follows directly. It is obvious that the equality holds when and only when Y is a constant multiple of X. ■

Remark If we apply Cauchy-Schwarz to two random variables $X - \mu_x$ and $Y - \mu_y$ centered around their means, then

$$\text{cov}(X, Y)^2 \;\leq\; \text{var}(X)\text{var}(Y)$$

It follows in particular that $|\text{cor}(X, Y)| \leq 1$ with equality when and only when Y is a linear function of X.

Recall that a set $A \subset \mathbf{R}^n$ is said to be *convex* if $\alpha x + (1-\alpha)y \in A$ for any $x, y \in A$ and $\alpha \in [0, 1]$. A function $f : \mathbf{R} \to \mathbf{R}$ is called *convex* if the set $\{(x, y)|\, y \geq f(x)\} \subset \mathbf{R}^2$ is convex.

Theorem 4 (Jensen) *Let $f : \mathbf{R} \to \mathbf{R}$ be convex. Then*

$$f(\mathbf{E}X) \leq \mathbf{E}f(X)$$

Proof Since $f : \mathbf{R} \to \mathbf{R}$ is convex, there exists a linear function $\ell : \mathbf{R} \to \mathbf{R}$ such that

$$\ell \leq f \quad \text{and} \quad \ell(\mathbf{E}X) = f(\mathbf{E}X)$$

It follows that

$$\begin{aligned} \mathbf{E}f(X) &\geq \mathbf{E}\ell(X) \\ &= \ell(\mathbf{E}X) \\ &= f(\mathbf{E}X) \end{aligned}$$

as was to be shown. ■

Remarks

(a) Functions such as $f(x) = |x|,\ x^2 and e^x$ are convex.

(b) We call a function $f : \mathbf{R} \to \mathbf{R}$ is *concave* if the set $\{(x, y) : y \leq f(x)\} \subset \mathbf{R}^2$ is convex. Obviously, the inequality is reversed for a concave function such as $f(x) = \log x$.

3. Conditional Expectation

Let X be a random variable with $\mathbf{E}(X) < \infty$, and $\mathcal{G}$ be a sub-σ-field. We define

Definition 1 *The conditional expectation* $\mathbf{E}(X|\mathcal{G})$ *of* X *given* $\mathcal{G}$ *is a random variable such that*

(a) $\mathbf{E}(X|\mathcal{G})$ *is* $\mathcal{G}$*-measurable,*

(b) *For every* $F \in \mathcal{G}$*,*

$$\int_F \mathbf{E}(X|\mathcal{G})\, d\mathbf{P} = \int_F X\, d\mathbf{P}$$

We also use the notation $\mathbf{E}(X|Y)$ to denote $\mathbf{E}(X|\sigma(Y))$.

Remark Let $\{F_k\}$ be a partition of Ω with $\mathbf{P}(F_k) > 0$, and suppose $\mathcal{G} = \sigma(\{F_k\})$ is the σ-field generated by $\{F_k\}$, i.e., $\mathcal{G}$ be the smallest σ-field including $\{F_k\}$. In this simple case,

(a) Condition (a) requires that $\mathbf{E}(X|\mathcal{G})$ must assign the same value to every outcome in each of F_k, i.e., it must be constant over each of the partition. Therefore, it must be given as

$$\mathbf{E}(X|\mathcal{G}) = \sum c_k \mathrm{I}(F_k)$$

with some constants c_k's.

(b) Condition (b) implies in particular that $c_k \mathbf{P}(F_k) = \int_{F_k} X \, d\mathbf{P}$. We thus have

$$c_k = \frac{1}{\mathbf{P}(F_k)} \int_{F_k} X \, d\mathbf{P}$$

which can be thought of the average of X over F_k with respect to the conditional probability of X. The conditional expectation $\mathbf{E}(X|\mathcal{G})$ may therefore be viewed as a random variable taking values that are local averages of X over the partitions made by $\mathcal{G}$.

Example Let E and F be two events such that

$$\mathbf{P}(E) = \mathbf{P}(F) = \frac{1}{2} \quad \text{and} \quad \mathbf{P}(E \cap F) = \frac{1}{3}$$

Define $X = \mathrm{I}(E), Y = \mathrm{I}(F)$ and let $\mathcal{G} = \sigma(Y) = \{\emptyset, \Omega, F, F^c\}$. Over F, the conditional distribution of X is given by probabilities $2/3$ and $1/3$ at points 1 and 0, respectively. The average value of X over F is therefore $2/3$. By the same token, the conditional distribution of X over F^c is given by the probabilities $1/3$ and $2/3$ on points 1 and 0. The local average of X over F^c is therefore $1/3$. The conditional expectation $\mathbf{E}(X|\mathcal{G})$ is therefore given by

$$\mathbf{E}(X|\mathcal{G}) = \frac{2}{3}\,\mathrm{I}(F) + \frac{1}{3}\,\mathrm{I}(F^c)$$

Existence

(a) Let $\mathbf{E}(X^2) < \infty$. In this case, we may view X as an element of a vector space of square integrable random variables defined on a common probability space, with inner product $\mathbf{E}(XY)$. Consider a subspace $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$ of $\mathcal{G}$-measurable random variables, and let Y be the *projection* of X onto this subspace. The projection theorem guarantees the existence of such a random variable. By construction, we have for all $W \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$,

$$\mathbf{E}(X - Y)W = 0$$

In particular, for $W = \mathrm{I}(F)$ for any $F \in \mathcal{G}$,

$$\mathbf{E}(X\,\mathrm{I}(F)) = \mathbf{E}(Y\,\mathrm{I}(F))$$

It therefore follows that $Y = \mathbf{E}(X|\mathcal{G})$, since the conditions for conditional expectation hold for Y. The conditional expectation $\mathbf{E}(X|\mathcal{G})$ can thus be interpreted as the *projection* of X onto a subspace in a vector space. Its existence is therefore given by the projection theorem.

(b) The existence of the conditional expectation is also guaranteed by the *Radon-Nikodym* theorem. To see this, notice that

$$\mu(F) = \int_F X\, d\mathbf{P} \quad \text{for} \quad F \in \mathcal{G}$$

defines a measure over $\mathcal{G}$. Since it is obviously absolutely continuous with respect to $\mathbf{P}$, the Radon-Nikodym theorem guarantees the existence of a $\mathcal{G}$-measurable random variable, say Y, such that for any $F \in \mathcal{G}$,

$$\mu(F) = \int_F Y\, d\mathbf{P}$$

Cleary, $Y = \mathbf{E}(X|\mathcal{G})$, since it satisfies the conditions for the conditional expectation.

Traditional Usage Suppose that X and Y are random variables with joint density $p(x, y)$. Let

$$\mathbf{E}(X|Y = y) = \int xp(x|y)\, d\mu(x)$$

which is traditionally defined as the conditional expectation of X given $Y = y$. The integral yields a function of y, which we denote by $f(y)$, i.e.,

$$f(y) = \mathbf{E}(X|Y = y)$$

For each $F \in \sigma(Y)$, there exists $B \in \mathcal{B}(\mathbf{R})$ such that $F = Y^{-1}(B)$. Therefore, we have

$$\begin{aligned}
\int_F f(Y)\, d\mathbf{P} &= \int_B f(y)p(y)\, d\mu(y) \\
&= \int_B \left(\int_{\mathbf{R}} xp(x|y)\, d\mu(x)\right) p(y)\, d\mu(y) \\
&= \iint_{\mathbf{R}\times B} xp(x, y)\, d\mu(x)\, d\mu(y) \\
&= \int_F X\, d\mathbf{P}
\end{aligned}$$

for all $F \in \sigma(Y)$. It follws that

$$f(Y) = \mathbf{E}(X|Y)$$

To get $\mathbf{E}(X|Y)$, we may just replace y with Y in $f(y)$.

Example Let

$$p(x,y) = (x+y)\,\mathrm{I}\{0 \le x, y \le 1\}$$

Then, for $0 \le y \le 1$,

$$\begin{aligned}\mathbf{E}(X|Y=y) &= \int xp(x|y)dx \\ &= \int_0^1 x \frac{x+y}{\frac{1}{2}+y} dx \\ &= \frac{\frac{1}{3}+\frac{y}{2}}{\frac{1}{2}+y}\end{aligned}$$

We now have

$$\mathbf{E}(X|Y) = \frac{\frac{1}{3}+\frac{Y}{2}}{\frac{1}{2}+Y} \quad \text{for} \quad 0 \le Y(\omega) \le 1$$

Properties We now list some of the useful properties of conditional expectation.

(a) *Law of Iterated Expectation* $\mathbf{E}(\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(X)$, which is straightforward from the definition (b) of conditional expectation with $F = \Omega$.

(b) If Y is $\mathcal{G}$-measurable, then $\mathbf{E}(XY|\mathcal{G}) = Y\mathbf{E}(X|\mathcal{G})$.

(c) *Linearity* $\mathbf{E}(\alpha X + \beta Y|\mathcal{G}) = \alpha\mathbf{E}(X|\mathcal{G}) + \beta\mathbf{E}(Y|\mathcal{G})$, which is obvious.

5. Miscellanies

Moments Let X be a random variable. We define the *k-th moment* μ_k and the *k-th central moment* μ_k^* respectively by

$$\mu_k = \mathbf{E}(X^k) \quad \text{and} \quad \mu_k^* = \mathbf{E}(X - \mathbf{E}(X))^k$$

Proposition 5

(a) *If* $\mu_q < \infty$, *then* $\mu_p < \infty$ *for all* $p \leq q$.

(b) $\mu_k < \infty$ *if and only if* $\mu_k^* < \infty$.

Proof Note that

$$\begin{aligned} \mathbf{E}|X|^p &\leq 1 + \mathbf{E}|X|^p \mathrm{I}\{|X| > 1\} \\ &\leq 1 + \mathbf{E}|X|^q \mathrm{I}\{|X| > 1\} \end{aligned}$$

for all $p \leq q$, which proves the part (a). The part (b) is immediate, upon noticing that μ_k^* is the expected value of a k-th polynomial in X. ■

For a random vector $X = (X_1, \ldots, X_n)'$, we define

$$\mathbf{E}(X) = \begin{pmatrix} \mathbf{E}(X_1) \\ \vdots \\ \mathbf{E}(X_n) \end{pmatrix} \quad \text{and} \quad \mathrm{var}(X) = \mathbf{E}(X - \mathbf{E}(X))(X - \mathbf{E}(X))'$$

Moment Generating Function Let X be a random variable with density p. The *moment generating function* of X is defined as

$$m(t) = \mathbf{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} p(x)\, d\mu(x)$$

Remarks

(a) The moment generating function is the *Laplace transform* of the density.

(b) One may easily see that

$$\frac{d^k}{dt^k} m(0) = \mathbf{E}(X^k)$$

whenever differentiation and integration are interchangeable, to which the name of moment generating function is due.

Characteristic Function The *characteristic function* of a random variable X is defined by

$$\varphi(t) = \mathbf{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} p(x) d\mu(x)$$

Remarks

(a) The characteristic function is the *Fourier transform* of the density.

(b) The moment generating function does not exist in some cases. However, one may easily see that the characteristic function always exists because

$$|e^{itx}| = 1$$

and hence bounded.

Quantile Let X be a random variable with distribution function F. We define the p-th *quantile* or *fractile* of X (or F) to be

$$F^{-1}(p) = \inf\{x \mid F(x) \geq p\}$$

for $0 < p < 1$. We sometimes call F^{-1} the inverse function of F. However, F^{-1} is *not* the usual inverse function of F, which does not exist in many cases. In particular, we call the 1/2-th quantile *median*, which is often denoted by $\mathrm{med}(X)$ for a random variable X. Note that F is right continuous, and min, instead of inf, can be used in the definition of the quantile.

6. Exercises

1. Let the sample space $\Omega = [0,1]$ and the probability on Ω be given by the density

$$p(x) = 2x$$

over $[0,1]$. We define random variables X and Y by

$$X(\omega) = \begin{cases} 1, & 0 \leq \omega < 1/4 \\ 0, & 1/4 \leq \omega < 1/2 \\ -1, & 1/2 \leq \omega < 3/4 \\ 0, & 3/4 \leq \omega \leq 1 \end{cases} \quad \text{and} \quad Y(\omega) = \begin{cases} 1, & 0 \leq \omega < 1/2 \\ 0, & 1/2 \leq \omega \leq 1 \end{cases}$$

(a) Find the conditional expectation $\mathbf{E}(X^2|Y)$.

(b) Show that $\mathbf{E}(\mathbf{E}(X^2|Y)) = \mathbf{E}(X^2)$.

2. Let the sample space $\Omega = \mathbf{R}$ and the probability $\mathbf{P}$ on Ω be given by

$$\mathbf{P}\left\{\frac{1}{3}\right\} = \frac{1}{3} \quad \text{and} \quad \mathbf{P}\left\{\frac{2}{3}\right\} = \frac{2}{3}$$

Define a sequence of random variables by

$$X_n = \left(3 - \frac{1}{n}\right)\ \mathrm{I}(A_n) \quad \text{and} \quad X = 3\ \mathrm{I}\left(\lim_{n\to\infty} A_n\right)$$

where

$$A_n = \left[\frac{1}{3} + \frac{1}{n},\ \frac{2}{3} + \frac{1}{n}\right)$$

for $n = 1, 2, \ldots$.

(a) Show that $\lim_{n\to\infty} A_n$ exists so that X is well defined.

(b) Compare $\lim_{n\to\infty} \mathbf{E}(X_n)$ with $\mathbf{E}(X)$.

(c) Prove or disprove that $\lim_{n\to\infty} \mathbf{E}(X_n - X)^2 = 0$.

3. The joint probability density function of X and Y is given by

$$p(x, y) = 3(x + y)\ \mathrm{I}\{0 \le x + y \le 1, 0 \le x, y \le 1\}$$

(a) Find $\mathbf{E}(Y|X)$.

(b) Find $\mathrm{cov}(X, Y)$.

4. Let X and Y have the joint density given by

$$p(x, y) = 8xy\ \mathrm{I}\{0 \le x \le y \le 1\}$$

(a) Compute $\mathbf{E}(X)$, $\mathbf{E}(Y)$ and $\mathbf{E}\left(\frac{X}{Y}\right)$.

(b) Notice that

$$\mathbf{E}\left(\frac{X}{Y}\right) = \frac{\mathbf{E}(X)}{\mathbf{E}(Y)}$$

Is this true in general? If not, explain why this holds here.

5. Show that

$$\mathbf{E}|X|^p \le (\mathbf{E}|X|^q)^{p/q}$$

for all $p \le q$.

6. Define

$$Y = \mathrm{I}\{X \ge 0\}$$

Show that

$$\begin{aligned} \mathrm{med}\,(Y) = 1 \quad &\text{iff} \quad \mathbf{P}\{Y = 1\} > 1/2 \\ &\text{iff} \quad \mathbf{P}\{X \geq 0\} > 1/2 \\ &\text{iff} \quad \mathrm{med}\,(X) \geq 0 \end{aligned}$$

to deduce that $\mathrm{med}\,(Y) = \mathrm{I}\,\{\mathrm{med}\,(X) \geq 0\}$.