Expected Value

Undergraduate definition of E: integral for absolutely continuous X, sum for discrete. But: $\exists$ rvs which are neither absolutely continuous nor discrete.

General definition of E.

A random variable X is **simple** if we can write

$$X(\omega) = \sum_{1}^{n} a_i 1(\omega \in A_i)$$

for some constants $a_1, \ldots, a_n$ and events A_i.

Def'n: For a simple rv X we define

$$E(X) = \sum a_i P(A_i)$$

For positive random variables which are not simple we extend our definition by approximation:

Def'n: If $X \geq 0$ (almost surely, $P(X \geq 0) = 1$) then

$$E(X) = \text{sup}\{E(Y) : 0 \leq Y \leq X, Y \text{ simple}\}$$

Def'n: We call X **integrable** if

$$E(|X|) < \infty\,.$$

In this case we define

$$E(X) = E(\text{max}(X, 0)) - E(\text{max}(-X, 0))$$

Facts: E is a linear, monotone, positive operator:

1. **Linear**: $E(aX+bY) = aE(X)+bE(Y)$ provided X and Y are integrable.

2. **Positive**: $P(X \ge 0) = 1$ implies $E(X) \ge 0$.

3. **Monotone**: $P(X \ge Y) = 1$ and X, Y integrable implies $E(X) \ge E(Y)$.

Major technical theorems:

Monotone Convergence: If $0 \le X_1 \le X_2 \le \cdots$ a.s. and $X = \lim X_n$ (which exists a.s.) then

$$E(X) = \lim_{n\to\infty} E(X_n)$$

Dominated Convergence: If $|X_n| \le Y_n$ and $\exists$ rv X st $X_n \to X$ a.s. and rv Y st $Y_n \to Y$ with $E(Y_n) \to E(Y) < \infty$ then

$$E(X_n) \to E(X)$$

Often used with all Y_n the same rv Y.

Fatou's Lemma: If $X_n \ge 0$ then

$$E(\liminf X_n) \le \liminf E(X_n)$$

Theorem: With this definition of E if X has density $f(x)$ (even in $\mathbb{R}^p$ say) and $Y = g(X)$ then

$$E(Y) = \int g(x)f(x)dx\,.$$

(This could be a multiple integral.)

Works even if X has density but Y doesn't.

If X has pmf f then

$$E(Y) = \sum_x g(x)f(x)\,.$$

Def'n: r^{th} moment (about origin) of a real rv X is $\mu_r' = E(X^r)$ (provided it exists). Generally use μ for $E(X)$. The r^{th} central moment is

$$\mu_r = E[(X-\mu)^r]$$

Call $\sigma^2 = \mu_2$ the variance.

Def'n: For an $\mathbb{R}^p$ valued rv X $\mu_X = E(X)$ is the vector whose i^{th} entry is $E(X_i)$ (provided all entries exist).

Def'n: The $(p \times p)$ variance covariance matrix of X is

$$Var(X) = E\left[(X-\mu)(X-\mu)^t\right]$$

which exists provided each component X_i has a finite second moment. More generally if $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ both have all components with finite second moments then

$$\text{Cov}(X,Y) = \text{E}\left[(X-\mu_X)(Y-\mu_Y)^T\right]$$

We have

$$\text{Cov}(AX+a, BY+b) = A\text{Cov}(X,Y)B^T$$

for general (conforming) matrices A, B and vectors a and b.

Moments and probabilities of rare events are closely connected as will be seen in a number of important probability theorems. Here is one version of Markov's inequality (one case is Chebyshev's inequality):

$$
\begin{aligned}
P(|X-\mu| \geq t) &= E[1(|X-\mu| \geq t)] \\
&\leq E\left[\frac{|X-\mu|^r}{t^r} 1(|X-\mu| \geq t)\right] \\
&\leq \frac{E[|X-\mu|^r]}{t^r}
\end{aligned}
$$

The intuition is that if moments are small then large deviations from average are unlikely.

Moments and independence

Theorem: If $X_1, \ldots, X_p$ are independent and each X_i is integrable then $X = X_1 \cdots X_p$ is integrable and

$$E(X_1 \cdots X_p) = E(X_1) \cdots E(X_p)$$

Proof: Usual order: simple Xs first, then positive, then integrable.

Suppose each X_i is simple:

$$X_i = \sum_j x_{ij} 1(X_i = x_{ij})$$

where the x_{ij} are the possible values of X_i.

Then

$$
\begin{aligned}
E(X_1 \cdots X_p) & \\
&= \sum_{j_1 \cdots j_p} x_{1j_1} \cdots x_{pj_p} \times \\
&\qquad E(1(X_1 = x_{1j_1}) \cdots 1(X_p = x_{pj_p})) \\
&= \sum_{j_1 \cdots j_p} x_{1j_1} \cdots x_{pj_p} \times \\
&\qquad P(X_1 = x_{1j_1} \cdots X_p = x_{pj_p}) \\
&= \sum_{j_1 \cdots j_p} x_{1j_1} \cdots x_{pj_p} \times \\
&\qquad P(X_1 = x_{1j_1}) \cdots P(X_p = x_{pj_p}) \\
&= \left[\sum_{j_1} x_{1j_1} P(X_1 = x_{1j_1}) \right] \times \cdots \times \\
&\qquad \left[\sum_{j_p} x_{pj_p} P(X_p = x_{pj_p}) \right] \\
&= \prod E(X_i)
\end{aligned}
$$

General $X_i \ge 0$: $X_{i,n}$ is X_i rounded down to the nearest multiple of 2^{-n} (to a maximum of n). Each $X_{i,n}$ is simple and $X_{1,n}, \ldots, X_{p,n}$ are independent. Thus

$$\text{E}(\prod X_{j,n}) = \prod \text{E}(X_{j,n})$$

for each n. If

$$X_n^* = \prod X_{j,n}$$

then

$$0 \le X_1^* \le X_2^* \le \cdots$$

and X_n^* converges to $X^* = \prod X_i$ so that

$$\text{E}(X^*) = \lim \text{E}(X_n^*)$$

by monotone convergence. Also by monotone convergence

$$\lim \prod \text{E}(X_{j,n}) = \prod \text{E}(X_j)) < \infty$$

This shows both that X^* is integrable and that

$$E(\prod X_j) = \prod \mathsf{E}(X_j)$$

The general case uses the fact that we can write each X_i as the difference of its positive and negative parts:

$$X_i = \max(X_i, 0) - \max(-X_i, 0)$$

Just expand out the product and use the previous case.

Lebesgue Integration

Lebesgue integral defined much the same way as E.

Borel function f *simple* if

$$f(x) = \sum_1^n a_i 1(x \in A_i)$$

for almost all $x \in \mathbb{R}^p$ and some constants a_i and Borel sets A_i with $\lambda(A_i) < \infty$). For such an f we define

$$\int f(x)dx = \sum a_i \lambda(A_i)$$

Again if

$$\sum a_i 1_{A_i} = \sum b_j 1_{B_j}$$

almost everywhere and all A_i and B_j have finite Lebesgue measure you must check that

$$\sum a_i \lambda(A_i) = \sum b_j \lambda(B_j)$$

If $f \ge 0$ almost everywhere and f is Borel define

$$\int f(x)dx = \text{sup}\{\int g(y)dy\}$$

where the sup ranges over all simple functions g such that $0 \le g(x) \le f(x)$ for almost all x. Call $f \ge 0$ integrable if $\int f(x)dx < \infty$.

Call a general f integrable if $|f|$ is integrable and define for integrable f

$$\int f(x)dx = \int \text{max}(f(x), 0)dx - \int \text{max}(-f(x), 0)dx$$

Remark: Again you must check that you have not changed the definition of f for either of the previous categories of f.

Facts: $\int$ is a linear, monotone, positive operator:

1. **Linear**: provided f and g are integrable
$$\int af(x)+bg(x)dx = a\int f(x)dx+b\,\mathrm{inf}\, g(x)dx$$

2. **Positive**: If $f(x) \geq 0$ almost everywhere then $\int f(x)dx \geq 0$.

3. **Monotone**: If $f(x) > g(x)$ almost everywhere and f and g are integrable then
$$\int f(x)dx \geq \int g(x)dx.$$

Each of these facts is proved first for simple functions then for positive functions then for general integrable functions.

Major technical theorems:

Monotone Convergence: If $0 \le f_1 \le f_2 \le \cdots$ almost everywhere and $f = \lim f_n$ (which has to exist almost everywhere) then

$$\int f(x)dx = \lim_{n\to\infty} f_n(x)dx$$

Dominated Convergence: If:

1) $|f_n| \le g_n$

2) there is a Borel function f such that $f_n(x) \to f(x)$ for almost all x

3) there is a Borel function g such that $g_n(x) \to g(x)$ with $\int g_n(x)dx \to \int g(x)dx < \infty$

Then f is integrable and

$$\int f_n(x)dx \to \int f(x)dx$$

Fatou's Lemma: If $f_n \ge 0$ almost everywhere then

$$\int \liminf f_n(x)dx \le \liminf \int f_n(x)dx.$$

Notice frequent use of almost all or almost everywhere in hypotheses. In def' of E wherever we require a property of the function $X(\omega)$ we can require it to hold only for a set of ω whose complement has probability 0. In this case we say the property holds **almost surely**. For instance the dominated convergence theorem is usually written:

Dominated Convergence: If $|X_n| \le Y_n$ almost surely (often abbreviated a.s.) and there is a random variable X such that $X_n \to X$ a.s. and a random variable Y such that $Y_n \to Y$ almost surely with $E(Y_n) \to E(Y) < \infty$ then

$$E(X_n) \to E(X)$$

Hypothesis of almost sure convergence can be weakened.

Multiple Integration: Lebesgue integrals over $\mathbb{R}^p$ defined using Lebesgue measure on $\mathbb{R}^p$.

Iterated integrals wrt Lebesgue measure on $\mathbb{R}^1$ give same answer.

Theorem[Tonelli]: If $f : \mathbb{R}^{p+q} \mapsto \mathbb{R}$ is Borel and $f \geq 0$ almost everywhere then for almost every $x \in \mathbb{R}^p$ the integral

$$g(x) \equiv \int f(x,y)dy$$

exists and

$$\int g(x)dx = \int f(x,y)dxdy$$

RHS denotes $p+q$ dimensional integral defined previously.

Theorem[Fubini] If $f : \mathbb{R}^{p+q} \mapsto \mathbb{R}$ is Borel and integrable then for almost every $x \in \mathbb{R}^p$ the integral

$$g(x) \equiv \int f(x,y)dy$$

exists and is finite. Moreover g is integrable and

$$\int g(x)dx = \int f(x,y)dxdy\,.$$

Results true for measures other than Lebesgue.