STAT 801=830 Expectation

Richard Lockhart

Simon Fraser University

STAT 801=830 — Fall 2012



What I assume you already know

- Continuous case: $E(X) = \int_{-\infty}^{\infty} xf(x) dx$.
- Discrete case: $E(X) = \sum_{x} x f(x)$
- Mean, variance, standard deviation, covariance, correlation.
- Conditional analogues of above.



What I want you to learn

- Abstract definition of E(X).
- Dominated, monotone convergence theorems.
- Mean, variance, standard deviation, covariance, correlation.
- Define conditional expectation and relation to conditional density.
- Give some properties of conditional expectation.



- Two elementary definitions of expected values:
- **Def'n**: If X has density f then

$$E\{g(X)\} = \int g(x)f(x) dx.$$

• **Def'n**: If X has discrete density f then

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

• FACT: if Y = g(X) for a smooth g

$$E(Y) = \int y f_Y(y) dy = \int g(x) f_Y(g(x)) g'(x) dx$$
$$= E\{g(X)\}$$

by change of variables formula for integration.

• Good: otherwise might have two different values for $E(e^X)$.



General Definition of E

p58, Appendix

- There are random variables which are neither absolutely continuous nor discrete.
- **Def'n**: RV X is simple if we can write

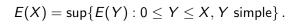
$$X(\omega) = \sum_{1}^{n} a_{i} 1(\omega \in A_{i})$$

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

• **Def'n**: For a simple rv X define

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

- For positive random variables which are not simple extend definition by approximation:
- **Def'n**: If X > 0 then





Integrable rvs

• **Def'n**: X is **integrable** if

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

In this case we define

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

- 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 $E(X) \ge 0$.
 - **Monotone**: $P(X \ge Y) = 1$ and X, Y integrable implies $E(X) \ge E(Y)$.



Convergence Theorems

- Major technical theorems:
- Monotone Convergence: If $0 \le X_1 \le X_2 \le \cdots$ and $X = \lim X_n$ (which has to exist) then

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

• **Dominated Convergence**: If $|X_n| \leq Y_n$ and \exists rv X such that $X_n \to X$ (technical details of this convergence later in the course) and a random variable Y such that $Y_n \to Y$ with $E(Y_n) \to E(Y) < \infty$ then

$$E(X_n) \rightarrow E(X)$$
.

- Often used with all Y_n the same rv Y.
- These theorems are used in approximation.



Connection to integration

With this definition of E:

① if X has density f(x) (even in \mathbb{R}^p say) and Y = g(X) then

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

(Could be a multiple integral.)

If X has pmf f then

$$E(Y) = \sum_{x} g(x)f(x).$$

First conclusion works, e.g., even if X has a density but Y doesn't.



Moments

pp 49-54

- **Def'n**: The r^{th} moment (about the origin) of a real rv X is $\mu'_r = E(X^r)$ (provided it exists).
- We generally use μ for E(X).
- **Def'n**: The r^{th} central moment is

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

- We call $\sigma^2 = \mu_2$ the variance.
- **Def'n**: For an \mathbb{R}^p valued random vector 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

$$Var(X) = E[(X - \mu)(X - \mu)^{t}]$$

which exists provided each component X_i has a finite second mome

Moments and independence

If X_1, \dots, 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

Suppose each X_i is simple:

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

where the x_{ii} are the possible values of X_i . Then

$$E(X_{1} \cdots X_{p}) = \sum_{j_{1} \dots j_{p}} x_{1j_{1}} \cdots x_{pj_{p}} E(1(X_{1} = x_{1j_{1}}) \cdots 1(X_{p} = x_{pj_{p}}))$$

$$= \sum_{j_{1} \dots j_{p}} x_{1j_{1}} \cdots x_{pj_{p}} P(X_{1} = x_{1j_{1}} \cdots X_{p} = x_{pj_{p}})$$

$$= \sum_{j_{1} \dots j_{p}} x_{1j_{1}} \cdots x_{pj_{p}} P(X_{1} = x_{1j_{1}}) \cdots P(X_{p} = x_{pj_{p}})$$

$$= \sum_{j_{1}} x_{1j_{1}} P(X_{1} = x_{1j_{1}}) \cdots \sum_{j_{p}} x_{pj_{p}} P(X_{p} = x_{pj_{p}})$$

$$= \prod_{j_{1} \dots j_{p}} E(X_{j_{1}}).$$



General Case

- General $X_i \geq 0$:
- Let X_{in} be X_i rounded down to nearest multiple of 2^{-n} (to maximum of n).
- That is: if

$$\frac{k}{2^n} \le X_i < \frac{k+1}{2^n}$$

then $X_{in} = k/2^n$ for $k = 0, ..., n2^n$.

- For $X_i > n$ put $X_{in} = n$.
- Apply case just done:

$$E(\prod X_{in}) = \prod E(X_{in}).$$

- Monotone convergence applies to both sides.
- General case: write each X_i as difference of positive and negative parts:

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

Apply positive case.



- Abstract definition of conditional expectation is:
- **Def'n**: E(Y|X) is any function of X such that

$$E[R(X)E(Y|X)] = E[R(X)Y]$$

for any bounded function R(X).

• **Def'n**: E(Y|X=x) is a function g(x) such that

$$g(X) = E(Y|X)$$

• Fact: If X, Y has joint density $f_{X,Y}(x,y)$ and conditional density f(y|x) then

$$g(x) = \int y f(y|x) dy$$

satisfies these definitions.



Proof

$$E(R(X)g(X)) = \int R(x)g(x)f_X(x)dx$$

$$= \int R(x) \int yf(y|x)dyf_X(x)dx$$

$$= \int \int R(x)yf_X(x)f(y|x)dydx$$

$$= \int \int R(x)yf_{X,Y}(x,y)dydx$$

$$= E(R(X)Y)$$



Interpretation of conditional expectation

- **Intution**: Think of E(Y|X) as average Y holding X fixed.
- Behaves like ordinary expected value but functions of X only are like constants:

$$E(\sum A_i(X)Y_i|X) = \sum A_i(X)E(Y_i|X)$$

• Statement called Adam's law by Jerzy Neyman - he used to say it comes before all the others:

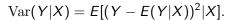
$$E[E(Y|X)] = E(Y)$$

which is just the definition of E(Y|X) with $R(X) \equiv 1$.

• In regression courses we say that the total sum of squares is the sum of the regression sum of squares plus the residual sum of squares:

$$Var(Y) = Var(E(Y|X)) + E[Var(Y|X)]$$

The conditional variance means



STAT 801=830 Expectation

