STAT 830
Expectation

Richard Lockhart

Simon Fraser University

STAT 830 — Fall 2013

What I assume you already know

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

What I want you to learn

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

Elementary definitions p 47-48

- Two elementary definitions of expected values:
- **Definition**: If X has density f then

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

- **Definition**: 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

$$\begin{aligned} E(Y) &= \int y f_Y(y)\,dy = \int g(x) f_Y(g(x)) g'(x)\,dx \\ &= E\{g(X)\} \end{aligned}$$

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.
- **Definition**: 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.

- **Definition**: 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:
- **Definition**: If $X \geq 0$ then

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

Integrable rvs

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

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 $|X_n| \le Y_n$ 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

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

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

Connection to integration

Theorem

With this definition of E:

1. *if X has density $f(x)$ (even in R^p say) and $Y = g(X)$ then*

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

   (Could be a multiple integral.)

2. *If X has pmf f then*

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

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

- **Definition**: 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 $E(X)$.
- **Definition**: The r^{th} central moment is

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

- We call $\sigma^2 = \mu_2$ the variance.
- **Definition**: For an 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).
- **Definition**: The $(p \times p)$ variance covariance matrix of X is

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

  which exists provided each component X_i has a finite second moment.

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

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

$$\begin{aligned}
E(X_1 \cdots X_p) &= \sum_{j_1 \ldots j_p} x_{1j_1} \cdots x_{pj_p} E(1(X_1 = x_{1j_1}) \cdots 1(X_p = x_{pj_p})) \\
&= \sum_{j_1 \ldots j_p} x_{1j_1} \cdots x_{pj_p} P(X_1 = x_{1j_1} \cdots X_p = x_{pj_p}) \\
&= \sum_{j_1 \ldots j_p} x_{1j_1} \cdots x_{pj_p} P(X_1 = x_{1j_1}) \cdots P(X_p = x_{pj_p}) \\
&= \sum_{j_1} x_{1j_1} P(X_1 = x_{1j_1}) \cdots \sum_{j_p} x_{pj_p} P(X_p = x_{pj_p}) \\
&= \prod E(X_i) .
\end{aligned}$$

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

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

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

Conditional Expectations pp 54-56

- Abstract definition of conditional expectation is:
- **Definition**: $E(Y|X)$ is any function of X such that

$$E\left[R(X)E(Y|X)\right] = E\left[R(X)Y\right]$$

  for any bounded function $R(X)$.
- **Definition**: $E(Y|X=x)$ 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

$$
\begin{aligned}
E(R(X)g(X)) &= \int R(x)g(x)f_X(x)dx \\
&= \int R(x) \int yf(y|x)dyf_X(x)dx \\
&= \int \int R(x)yf_X(x)f(y|x)dydx \\
&= \int \int R(x)yf_{X,Y}(x,y)dydx \\
&= E(R(X)Y)
\end{aligned}
$$

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:
$$\mathrm{Var}(\mathrm{Y}) = \mathrm{Var}(E(Y|X)) + E[\mathrm{Var}(Y|X)]$$
- The conditional variance means
$$\mathrm{Var}(Y|X) = E[(Y - E(Y|X))^2|X].$$