Parametric Statistics

Conditional Expectation, Moments

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

October 16, 2023

Last Time

- Expectation is a summary of a distribution.
- We can compute the expectation of a function of an RV using LOTUS.
- Properties of expectation.
- Variance is a summary of how spread out a distribution is.
- Covariance describes how much two variables vary together.
- Correlation is covariance without scale.

Lecture Summary

4.7 Conditional Expectations

4.4 Moments

Conditional Expectation

Let X and Y be random variables such that the mean of Y exists and is finite. The conditional expectation (or conditional mean) of Y given $X = x$ is denoted by $E(Y \mid x)$ and is defined to be the expectation of the conditional distribution of Y given $X = x$. For example, if Y has a continuous conditional distribution given $X = x$ with conditional p.d.f. $g_2(y \mid x)$, then

$$E(Y \mid x) = \int_{-\infty}^{\infty} y g_2(y \mid x) dy.$$

Similarly, if Y has a discrete conditional distribution given $X = x$ with conditional p.f. $g_2(y \mid x)$, then

$$E(Y \mid x) = \sum_{\text{All } y} y g_2(y \mid x).$$

Conditional Expectation

- Conditional distributions are distributions, so they have expectations and variances.
- $E(Y|X = x) = \sum_y yP(y|X = x)$ is the conditional expectation of Y if you know $X = x$ (a number).
- $E(Y|X) = h(X)$ is a function of X. For every possible value x of X, $E(Y|X)$ takes the value $E(Y|X = x)$. So $E(Y|X)$ is a random variable.
- Law of total expectation/Law of iterated expectations:

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

Conditional Variance

Definition

For every given value x, let $\mathrm{Var}(Y \mid x)$ denote the variance of the conditional distribution of Y given that $X = x$. That is,

$$\mathrm{Var}(Y \mid x) = E\left\{[Y - E(Y \mid x)]^2 \mid x\right\}.$$

We call $\mathrm{Var}(Y \mid x)$ the conditional variance of Y given $X = x$.

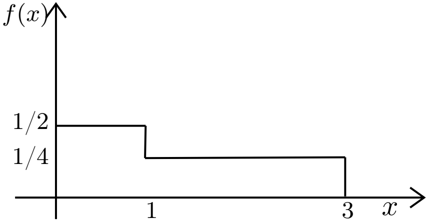
Law of total variance

If X and Y are arbitrary random variables for which the necessary expectations and variances exist, then

$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}[E(Y \mid X)]$$

We call $\mathrm{Var}(Y \mid x)$ the conditional variance of Y given $X = x$.

Using the laws of total expectation and variance



- $Y \sim Bernoulli(0.5)$.
- $X|Y = 0 \sim Uniform([0, 1])$.
- $X|Y = 1 \sim Uniform([1, 3])$
- Find $E(X|Y), Var(X|Y)$.
- Find $E(X), Var(X)$.

Moments and Central Moments

Definition (Moments and Central Moments)

Let X be a random variable and k be a positive integer.
The expectation $E(X^k)$ is the k-th moment of X.
The expectation $E[(X - E(X))^k]$ is the k-th central moment of X.

- The first moment is the mean: $\mu = E(X^1)$.
- The first central moment is zero: $E[(X - E(X))^1] = E(X - \mu) = E(X) - E(X) = 0$
- The second central moment is the variance: $E[(X - E(X))^2] = Var(X)$

Moments and Central Moments

- The k th moment exists if and only if $E\left(|X|^k\right) < \infty$.
- If the random variable X is bounded $(\Pr(a \leq X \leq b) = 1)$, then all moments of X must necessarily exist.
- It is possible, however, that all moments of X exist even though X is not bounded.
- If $E\left(|X|^k\right) < \infty$ for some positive integer k, then $E\left(|X|^j\right) < \infty$ for every positive integer j such that $j < k$.
- If the distribution of X is symmetric with respect to its mean μ, and if the central moment $E\left[(X-\mu)^k\right]$ exists for a given odd integer k, then the value of $E\left[(X-\mu)^k\right]$ will be 0.

Skewness

Let X be a random variable with mean μ, standard deviation σ, and finite third moment. The skewness of X is defined to be $E\left[(X-\mu)^3\right]/\sigma^3$.

- The reason for dividing the third central moment by σ^3 is to make the skewness measure only the lack of symmetry rather than the spread of the distribution.

Skewness

Let X be a random variable with mean μ, standard deviation σ, and finite third moment. The skewness of X is defined to be $E\left[(X-\mu)^3\right]/\sigma^3$.

- The reason for dividing the third central moment by σ^3 is to make the skewness measure only the lack of symmetry rather than the spread of the distribution.
- Let's compute the skewness of the Bernoulli distribution.
  - with $p = 0.5$
  - with $p = 0.1$

Moment Generating Functions

Definition

Let X be a random variable. The function

$$\psi(t) = E(e^{tX}), t \in R$$

is called the moment generating function (m.g.f.) of X.

Let X be a random variables whose m.g.f. $\psi(t)$ is finite for t in an open interval around zero. Then the $n-th$ moment of X is finite, for $n = 1, 2, \ldots,$ and

$$E(X^n) = \frac{d^n \psi(t)}{dt^n}|_{t=0}$$

Properties of Moment Generating Functions

- $\psi(aX + b) = e^{bt}\psi_X(at)$.
- Let $Y = \sum_{i=1}^{n} X_i$ where $X_1, \ldots, X_n$ are independent random variables with m.g.f $\psi_i(t)$ for $i = 1, \ldots, n$. Then

$$\psi_Y(t) = \prod_{i=1}^{n} \psi_i(t)$$

- Let X and Y be two random variables with m.g.f.'s $\psi_X(t)$ and $\psi_Y(t)$. If the m.g.f.'s are finite and $\psi_X(t) = \psi_Y(t)$ for all values of t in an open interval around zero, then X and Y have the same distribution.

Finding the p.f.'s for sums of random variables

- Find the mgf of Binomial(n,p).

Sum of Binomials is a Binomial

If X_1 and X_2 are independent random variables, and if X_i has the binomial distribution with parameters n_i and p $(i = 1, 2)$, then $X_1 + X_2$ has the binomial distribution with parameters $n_1 + n_2$ and p.

Mean and Median

Median

Let X be a random variable. Every number m with the following property is called a median of the distribution of X :

$$\Pr(X \leq m) \geq 1/2 \quad \text{and} \quad \Pr(X \geq m) \geq 1/2.$$

Mean Squared Error/M.S.E

The number $E\left[(X-d)^2\right]$ is called the mean squared error (M.S.E.) of the prediction d.

Mean Absolute Error/M.A.E.

The number $E(|X-d|)$ is called the mean absolute error (M.A.E.) of the prediction d.

Mean and Median

Mean minimizes M.S.E.

Let X be a random variable with finite variance σ^2, and let $\mu = E(X)$. For every number d,

$$E\left[(X-\mu)^2\right] \le E\left[(X-d)^2\right]$$

Median minimizes M.A.E.

Let X be a random variable with finite mean, and let m be a median of the distribution of X. For every number d,

$$E(|X-m|) \le E(|X-d|).$$

Equality holds if and only if d is also a median of the distribution of X.

Recap

- Conditional Expectation, Conditional Variance are functions of the conditioning variable.
- Law of total expectation/law of total variance can help us compute variances and expectations of complex functions.
- The means of powers X^k of an RV X are called moments of X.
- They can help us derive distributions of sums of independent random variables and prove limiting properties of distributions.