EE126: Probability and Random Processes

Lecture 14: Total Variance, Transforms

Abhay Parekh

UC Berkeley

March 8, 2011

Midterm

- It was not an easy exam. You did really well as a group! Most of you should feel very good about your performance.
- Regrades until Thursday. See your GSIs or me. I will make final call.
- Please look at the exam solutions.
- Let's wait until the last 10 mins to discuss more.

Convolution

Let $Z = X + Y$ and assume that X, Y continuous and independent.

$$\boxed{f_Z(z) = \int_x f_X(x) f_Y(z-x) = \int_y f_Y(y) f_X(z-y) dx}$$

Graphical Convolution: X, Y, Z uniform $[0,1]$, $W = X + Y + Z$.

Covariance

Given X, Y:

$$cov(X, Y) = E[XY] - E[X]E[Y] = E[(X - E[X])(Y - E[Y])]$$

- $cov(X, X) = var(X)$
- $cov(aX + b, Y) =$
- $cov(a, Y) = 0$
- $cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$
- Covariance $=0 \Rightarrow E[X] = E[X|Y]$
- X, Y independent $\Rightarrow$ covariance $=0$
- Covariance $= 0 \not\Rightarrow$, X, Y independent.

Correlation Coefficient

For any two random variables, X and Y with non zero variance, the correlation coefficient $\rho(X, Y)$ is

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}.$$

Special cases for the correlation coefficient, $\rho(X, Y)$:

$$\rho(X, Y) = \begin{cases} 1, & Y = aX + b \ \ a > 0; \\ -1, & Y = aX + b, \ a < 0; \\ 0, & E[X|Y] = E[X]. \end{cases}$$

More results

Sum of Variances

Given $X_1, ..., X_n$:

$$var(\sum_i X_n) = \sum_i var(X_i) + \sum_i \sum_{j \neq i} cov(X_i, X_j)$$

Iterated Expectation

Given two random variables, X, Y:

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

Estimating with X from Y: $E[X|Y]$

Suppose we want to estimate X but have no observations.
How to find the $\hat{X}$ which minimizes $E[(X-\hat{X})^2]$, i.e. the mean square error?

$$\begin{aligned} E[(X-\hat{X})^2] &= var(X-\hat{X}) + (E[X-g(X)])^2 \\ &= var(X) + (E[X-\hat{X}])^2 \\ &= var(X) + (E[X]-\hat{X})^2 \end{aligned}$$

So pick $\hat{X} = E[X]$ Now suppose we make an observation for random variable Y, i.e. $Y = y$. Then what should our estimate be? Again, we want to minimize mean square error (given $Y = y$) so:

$$E[(X-\hat{X})^2|Y=y] \text{ is minimized at } \hat{X} = E[X|Y=y]$$

$E[X|Y]$: Estimation Error

The mean of the estimate:

$$E[\hat{X}] = E[E[X|Y]] = E[X]$$

Also,

$$E[\underbrace{X - \hat{X}}_{\text{estimation error}}] = E[X - E[X|Y]] = E[X] - E[X] = 0$$

An estimator with zero average estimation error is called **unbiased**.

$E[X|Y]$ is an unbiased estimator of X.

Estimating with X from Y: $E[X|Y]$

$\hat{X}$ is uncorrelated with the estimation error $\hat{X} - X$.

$$\begin{aligned} cov(\hat{X}, \hat{X} - X) &= E[\hat{X}(\hat{X} - X)] - E[\hat{X}]E[\hat{X} - X] \\ &= E[\hat{X}(\hat{X} - X)] - E[X]0 \\ &= E[(\hat{X})^2] - E[X\hat{X}] \\ &= E[(\hat{X})^2] - E[E[X\hat{X}|Y]] \\ &= E[(\hat{X})^2] - E[(\hat{X})^2] \\ &= 0 \end{aligned}$$

So

$$var(\hat{X} + X - \hat{X}) = var(\hat{X}) + \mathrm{var}(X - \hat{X})$$

So

$$\boxed{var(X) = var(E[X|Y]) + var(X - E[X|Y])}$$

Law of Total Variance

Since $E[X - \hat{X}] = 0$,
$var(X - \hat{X}) = E[(X - \hat{X})^2] = E[E[(X - \hat{X})^2]|Y]$.
Now consider the random variable $X|Y$. Then

$$E[var(X|Y)] = E[E[(X - E[X|Y])^2]|Y].$$

In the previous slide we showed that:

$$\boxed{var(X) = var(E[X|Y]) + var(X - E[X|Y])}$$

Substituting:

Given random variables, X, Y:

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

Example: Bias of Coin

We toss a biased coin n times. Y: prob of heads, and X: number of heads. Y is distributed uniformly over $[0, 1]$. What are $E[X]$ and $var(X)$?

$$\hat{X} = E[X|Y] = nY$$

$$\boxed{E[X] = E[E[X|Y]] = E[nY] = \frac{n}{2}}$$

$$var(E[X|Y]) = var(nY) = n^2 var(Y) = \frac{n^2}{12}$$

$$var(X|Y) = nY(1 - Y)$$

$$E[var(X|Y)] = n(\frac{1}{2} - \frac{1}{3}) = \frac{n}{6}$$

$$\boxed{var(X) = \frac{n^2}{12} + \frac{n}{6}}$$

Example: Bias of a Coin Continued

Same problem: Let $X_i = 1$ if toss i is a head and $X_i = 0$ o.w.
What is $cov(X_i, X_j)$, $i \neq j$?

$$cov(X_i X_j) = E[X_i X_j] - E[X_i]E[X_j]$$

$$E[X_i] = E[E[X_i|Y]] = E[Y] = 0.5$$

$$E[X_i X_j] = E[E[X_i X_j|Y]] = E[E[X_i|Y]E[X_j|Y]] = E[Y^2] = \int_0^1 y^2 dy = \frac{1}{3}$$

$$cov(X_i X_j) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} = var(Y)$$

Therefore the tosses are not independent...
Check result for $var(X)$:

$$var(X_i) = E[X_i^2] - E[X_i]^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

$$var(X_1 + ... + X_n) = \frac{n}{4} + \frac{n(n-1)}{12} = \frac{1}{12}(3n + n^2 - n) = \frac{n^2}{12} + \frac{n}{6}$$

Summing a Random Number of Random Variables

Suppose $Y = X_1 + \ldots X_N$, the X_i are iid, but N is a random variable independent of the X_i's. What are $E[Y]$ and $var(Y)$?

$$E[Y] = E[E[Y|N]] = E[NE[X_i]] = E[N]E[X_i]$$

$$\boxed{E[Y] = E[N]E[X]}$$

$$var(Y) = var(E[Y|N]) + E[var(Y|N)$$

Now

$$var(E[Y|N]) = var(NE[X_i]) = E[X_i]^2 var(N)$$

$$E[var(Y|N)] = E[Nvar(X_i)] = var(X_i)E[N]$$

So

$$\boxed{var(Y) = E[X_i]^2 var(N) + E[N]var(X_i)}$$

Moment Generating Functions - Transforms

Sometimes rather than working with $f_X(x)$ we work with $E[e^{sx}]$ where s is any scalar. This is the **Transform or Moment Generating Function of** X.

Why?

1. It is easier to find $E[X^k]$, i.e. the moments of X (differentiate rather than integrate)
2. It is easier to add independent random variables (multiply rather than convolve)
3. It is easier prove things (e.g. Central Limit Theorem)

Given a random variable X, the Transform of X, $M_X(s)$ is defined as

$$M_X(s) = E[e^{sX}]$$

for all scalars s

Generating Moments with Transforms

Use the result that

$$e^{sx} = 1 + sx + \frac{s^2x^2}{2!} + \frac{s^3x^3}{3!} + \dots$$

Let X be a rv. Now use Linearity of Expectations:

$$E[e^{sx}] = 1 + sE[x] + \frac{s^2}{2!}E[X^2] + \dots$$

Now observe that

$$\begin{aligned} \frac{dE[e^{sx}]}{ds}\Big|_{s=0} &= E[X] \\ \frac{d^2E[e^{sx}]}{ds^2}\Big|_{s=0} &= E[X^2] \\ \frac{d^3E[e^{sx}]}{ds^3}\Big|_{s=0} &= E[X^3] \\ &\dots \end{aligned}$$

Moment Generating Function

For $M_X(s) = E[e^{sx}]$:

$$\frac{d^n M_X(s)}{ds^n}\Big|_{s=0} = E[X^n]$$

Properties:

1. $M_X(0) = 1$
2. If $X > 0$, $M_X(-\infty) = 0$ and if $X < 0$ then $M_X(\infty) = 0$.
3. If $Y = aX + b$ than
   $M_Y(s) = E[e^{s(aX+b)}] = e^{sb}E[e^{aX}] = e^{sb}M_X(as)$

Example: Exponential Distribution

$f_X(x) = \lambda e^{-\lambda x} \Rightarrow E[e^{sx}] = \lambda \int_{x=0}^{\infty} e^{sx} e^{-\lambda x} dx$

$$M_X(s) = \frac{\lambda}{\lambda - s}$$

$M(0) = 1, \lim_{s \to -\infty} M_X(s) = 0.$
Also, if $Y = aX + b$ then

$$M_Y(s) = e^{sb} M_X(as) = e^{sb} \frac{\lambda}{\lambda - as}$$

$$E[Y] = be^{sb} \frac{\lambda}{\lambda - as} + e^{sb} \frac{\lambda}{(\lambda - as)^2} a \Big|_{s=0}$$

$$E[Y] = b + \frac{a}{\lambda}$$

Inversion of Transform

It is somewhat surprising that a given transform corresponds to a unique CDF, i.e. $M_X(s)$ contains all the information in $f_X(x)$.
Why is this true? $M_X(s)$ is the bilateral Laplace transform of $f_X(x)$.
The inversions are usually done via pattern matching...
Example:

$$M_X(s) = \frac{1}{2}e^{-3s} + \frac{1}{4}e^{200s} + \frac{1}{4}e^{s}$$

$$p_X(x) = \begin{cases} -3, & \text{with prob } 0.5; \\ 200, & \text{with prob } 0.25; \\ 1, & \text{with prob } 0.25. \end{cases}$$

Helps to know $M_X(s)$ for popular distributions.
We won't require you to know $f_X(x), M_X(s)$ pairs.

Combinations

1. Mixture of distributions: Suppose $\sum_{i=1}^{n} p_i = 1$, and $f_X(x) = \sum_{i=1}^{n} p_i f_{X_i}(x)$. Then

$$\boxed{M_X(s) = \sum_{i=1}^{n} p_i M_{X_i}(s)}$$

2. Sum of Independent Random Variables: $Z = X + Y$; X, Y independent. Then

$$M_Z(s) = E[e^{(X+Y)s}] = E[e^{Xs}e^{Ys}] = E[e^{Xs}]E[e^{Ys}] = M_X(s)M_Y(s)$$

> So convolving the densities corresponds to multiplying transforms.

Example

If X_i is bernoulli with with parameter p then $M_{X_i} = 1 - p + pe^s$ for $i = 1, 2, ..., n$.
$Y = \sum_i X_i$ is a Binomial Random Variable.

$$M_Y(s) = \Pi_{i=1}^n (1 - p + pe^s) = (1 - p + pe^s)^n.$$

$$E[X] = n(1 - p + pe^s)^{n-1} pe^s \Big|_{s=0} = n(1)^{n-1} p = np$$

$$\begin{aligned} E[X^2] &= np[(n-1)(1 - p + pe^s)^{n-2} pe^{2s} + (1 - p + pe^s)^{n-1} e^s] \Big|_{s=0} \\ &= np(1 - p + np). \end{aligned}$$

Summing a Random Number of Random Variables

Let $Y = X_1 + ... + X_N$ where $X_i, i = 1, 2, ..., n$ are iid and N is a random variable.
Then $E[e^{sY}|N = n] = (M_X(s))^n$. Using Iterated Expectations:

$$M_Y(s) = E[e^{sY}] = E[E[e^{sY}|N = n]] = E[(M_X(s))^n]$$

Recall that $a^n = e^{\ln a}$:

$$(M_X(s))^n = e^{n \ln M_X(s)}$$

So

$$E[(M_X(s))^n] = \sum_{n=0}^{\infty} e^{n \ln(M_X(s))} p_N(n)$$

Now since

$$M_N(s) = \sum_{n=0}^{\infty} e^{sn} p_N(n)$$

$$M_Y(s) = E[(M_X(s))^n] = M_N(\ln M_X(s))$$

Transform of Sum of Random Number of RVs

To find $M_Y(s)$:

1. Find $M_N(s)$
2. Replace s with $\ln M_X(s)$, i.e. e^s with $M_X(s)$.

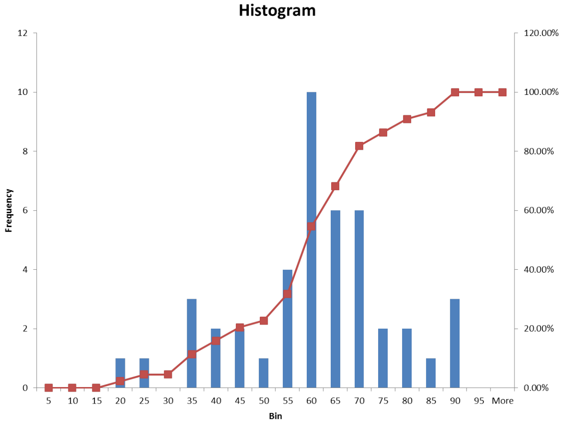
Example:
Each of 3 gas station is open on any given day with prob $\frac{1}{2}$
The amount of gas available is uniformly distributed on $[0, 1000]$.
Let Y be the total amount of gas available on any given day. Find $M_Y(s)$.
N: number of gas stations open:
$M_N(n) = (1 - 0.5 + 0.5e^s)^3 = \frac{1}{8}(1 + e^s)^3$.
Now

$$M_X(s) = \frac{e^{1000s} - 1}{1000s}$$

(Look this up)
So

$$M_Y(s) = \frac{1}{8}(1 + \frac{e^{1000s} - 1}{1000s})^3$$

Midterm



Your score

≤ 50: Concepts, App
$(50, 70]$: Concepts, App
> 70: Concepts, App.

Mean=58.41, Median = 60 Standard Deviation=16.39.