Lecture 15: Feb 9, Expectations, Variance and Covariance (Ross 7.1, 7.4)

15.1: Expectations of functions and sums

(i) For a discrete random variable, $\mathrm{E}(g(X)) \ = \ \sum_{x \in \mathcal{X}} g(x)\ p_X(x)$.

For a continuous random variable, $\mathrm{E}(g(X)) \ = \ \int_x g(x)\ f_X(x)\ dx$.

(This was proved for the discrete case in 394; Ross P.145)

If X and Y have joint pmf $p_{X,Y}(x,y)$ or pdf $f_{X,Y}(x,y)$ then $\mathrm{E}(g(X,Y)) \ = \ \sum_x \sum_y g(X,Y)\ p_{X,Y}(x,y)$ or $\mathrm{E}(g(X,Y)) \ = \ \int_x \int_y g(X,Y)\ f_{X,Y}(x,y)\ dx\ dy$ respectively. (This would be proved exact same way.)

(ii) Recall also (from 394), $\mathrm{E}(g_1(X) + g_2(X)) \ = \ \int (g_1(x) + g_2(x)) f_X(x)\ dx \ = \ \mathrm{E}(g_1(X)) + \mathrm{E}(g_2(X))$.

Now, if X and Y have joint pmf $p_{X,Y}(x,y)$ or pdf $f_{X,Y}(x,y)$ then

$$\begin{aligned}
\mathrm{E}(g_1(X) \ + \ g_2(Y)) \ &= \ \int_x \int_y (g_1(X) \ + \ g_2(Y))\ f_{X,Y}(x,y)\ dx\ dy \\
&= \ \int_x g_1(X) \left(\int_y f_{X,Y}(x,y) dy \right) dx \ + \ \int_y g_2(Y) \left(\int_x f_{X,Y}(x,y) dx \right) dy \\
&= \ \int_x g_1(X) f_X(x) dx \ + \ \int_y g_2(Y) f_Y(y) dy \ = \ \mathrm{E}(g_1(X)) \ + \ \mathrm{E}(g_2(Y))
\end{aligned}$$

15.2: Expectation of a product of (functions of) independent rvs

If X and Y are independent random variables

$$\begin{aligned}
\mathrm{E}(g_1(X) g_2(Y)) \ &= \ \int_x \int_y (g_1(X) g_2(Y))\ f_{X,Y}(x,y)\ dx\ dy \ = \ \int_x \int_y (g_1(X) g_2(Y))\ f_X(x) f_Y(y)\ dx\ dy \\
&= \ \left(\int_x g_1(X) f_X(x)\ dx \right) \left(\int_y g_2(Y) f_Y(y)\ dy \right) \ = \ \mathrm{E}(g_1(X)) \mathrm{E}(g_2(Y))
\end{aligned}$$

The proof for discrete random variables is similar.

15.3: Variance, Covariance, and correlation

(i) Recall, if $\mathrm{E}(X) = \mu$,

$\mathrm{var}(X) \ \equiv \ \mathrm{E}((X - \mu)^2) \ = \ \mathrm{E}(X^2 - 2\mu X + \mu^2) \ = \ \mathrm{E}(X^2) - 2\mu \mathrm{E}(X) + \mu^2 \ = \ \mathrm{E}(X^2) - (\mathrm{E}(X))^2$

(ii) If $\mathrm{E}(X) = \mu$ and $\mathrm{E}(Y) = \nu$, define $\mathrm{cov}(X,Y) \ \equiv \ \mathrm{E}((X - \mu)(Y - \nu))$.

Then $\mathrm{cov}(X,Y) \ = \ \mathrm{E}(XY - \mu Y - \nu X + \mu\nu) \ = \ \mathrm{E}(XY) - \mu \mathrm{E}(Y) - \nu \mathrm{E}(X) + \mu\nu \ = \ \mathrm{E}(XY) \ - \ \mathrm{E}(X)\mathrm{E}(Y)$.

Note $\mathrm{var}(X) \ = \ \mathrm{cov}(X,X)$, and $\mathrm{cov}(X, -Y) \ = \ -\mathrm{cov}(X,Y)$.

(iii) We see from **15.2** that if X and Y are independent, then $\mathrm{cov}(X,Y) = 0$.

(iv) The converse is NOT true: i.e. $\mathrm{cov}(X,Y) = 0$ does **not** imply X,Y independent.

Example: X and Y uniform on a circle/disc: $X = \cos(U)$, $Y = \sin(U)$, where $U \ \sim U(0, 2\pi)$.

(v) Define the *correlation coefficient* ρ by

$$\rho(X,Y) \ = \ \mathrm{cov}(X,Y) / \sqrt{\mathrm{var}(X)\mathrm{var}(Y)}$$

Note from (iii), if X and Y are independent, $\rho(X,Y) = 0$.

As in (iv), in general, the converse is NOT true.

Also note $\rho(X,X) = +1$, $\rho(X, -X) = \ -1$.

We shall show below that $-1 \leq \rho \leq 1$ always.

Lecture 16: Feb 11, Variances and covariances of sums of random variables (Ross 7.4)

16.1: Variance and covariance of a sum

(i) Let X_i have mean μ_i, $i = 1, ..., n$, and Y_j have mean ν_j, $j = 1, ..., m$. So $\mathrm{E}(\sum_1^n X_i) = \sum_1^n \mu_i$ and $\mathrm{E}(\sum_1^m Y_j) = \sum_1^m \nu_j$.

$$\begin{aligned}
\mathrm{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) &= \mathrm{E}\left(\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right)\left(\sum_{j=1}^n Y_j - \sum_{j=1}^m \nu_j\right)\right) = \mathrm{E}\left(\sum_{i=1}^n (X_i - \mu_i) \sum_{j=1}^m (Y_j - \mu_j)\right) \\
&= \mathrm{E}(\sum_{i=1}^n \sum_{j=1}^m (X_i - \mu_i)(Y_j - \nu_j)) = \sum_{i=1}^n \sum_{j=1}^m \mathrm{E}((X_i - \mu_i)(Y_j - \nu_j)) = \sum_{i=1}^n \sum_{j=1}^m \mathrm{cov}(X_i, Y_j).
\end{aligned}$$

(ii) $$\mathrm{var}\left(\sum_{i=1}^n X_i\right) = \mathrm{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n \sum_{j=1}^n \mathrm{cov}(X_i, X_j) = \sum_{i=1}^n \mathrm{var}(X_i) + 2 \sum\sum_{i<j} \mathrm{cov}(X_i, X_j)$$

(iii) If X_i and X_j are independent for all pairs (X_i, X_j), then $\mathrm{cov}(X_i, X_j) = 0$ so

$$\mathrm{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{var}(X_i)$$

16.2: The correlation inequality

Let X have variance σ_X^2 and Y have variance σ_Y^2.

$$0 \leq \mathrm{var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) = \frac{\mathrm{var}(X)}{\sigma_X^2} + \frac{\mathrm{var}(Y)}{\sigma_Y^2} \pm 2\frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = 2(1 \pm \rho(X,Y))$$

Hence $0 \leq (1 - \rho(X,Y))$ so $\rho \leq 1$; $0 \leq (1 + \rho(X,Y))$ so $\rho \geq -1$. i.e. $-1 \leq \rho \leq 1$.

16.3: Mean and variance of a sample mean Let $X_1, ..., X_n$ be *independent and identically distributed* (i.i.d.) each with mean μ and variance σ^2. The *sample mean* is defined as $\overline{X} = n^{-1} \sum_i X_i$. Then

$$\begin{aligned}
\mathrm{E}(\overline{X}) &= \mathrm{E}(n^{-1} \sum_{i=1}^n X_i) = n^{-1} \sum_{i=1}^n \mathrm{E}(X_i) = (n\mu)/n = \mu, \\
\mathrm{var}(\overline{X}) &= \mathrm{var}(n^{-1} \sum_{i=1}^n X_i) = n^{-2} \sum_{i=1}^n \mathrm{var}(X_i) = (n\sigma^2)/n^2 = \sigma^2/n
\end{aligned}$$

We can estimate μ by $\overline{X}$, and the variance of this estimator is σ^2/n: but now we need to estimate σ^2.

16.4: Mean of a sample variance Let $X_1, ..., X_n$ be i.i.d. each with mean μ and variance σ^2.

Note $\mathrm{E}(\sum_i (X_i - \mu)^2) = n\sigma^2$; but we usually do not know μ.

The *sample variance* is defined as $S^2 = \sum_i (X_i - \overline{X})^2/(n-1)$. Then

$$\begin{aligned}
(n-1)S^2 &\equiv \sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n (X_i - \mu + \mu - \overline{X})^2 \\
&= \sum_{i=1}^n \left((X_i - \mu)^2 - (X_i - \mu)(\overline{X} - \mu) + (\overline{X} - \mu)^2\right) \\
&= \sum_{i=1}^n (X_i - \mu)^2 - 2(\overline{X} - \mu) \sum_{i-1}^n (X_i - \mu) + n(\overline{X} - \mu)^2 \\
&= \sum_{i=1}^n (X_i - \mu)^2 - 2(\overline{X} - \mu)n(\overline{X} - \mu) + n(\overline{X} - \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\overline{X} - \mu)^2 \\
\mathrm{E}(S^2) &= (n-1)^{-1}\left(\sum_{i=1}^n \mathrm{E}((X_i - \mu)^2) - n\mathrm{E}((\overline{X} - \mu)^2)\right) \\
&= (n-1)^{-1}(n\mathrm{var}(X_i) - n\mathrm{var}(\overline{X})) = (n-1)^{-1}(n\sigma^2 - n(\sigma^2/n)) = \sigma^2
\end{aligned}$$

Lecture 17: Feb 13: Moment generating functions Ross 7.7

17.1: Definition and basic properties

(i) Definition: $M_X(t) = \mathrm{E}(e^{tX})$, provided expectation exists. Note $M_X(0) \equiv 1$.

Discrete case: $M_X(t) = \sum_x e^{tx} p_X(x)$. Continuous case: $M_X(t) = \int_{x=-\infty}^{\infty} e^{tx} f_X(x) dx$.

(ii) Moments: Differentiating: $M'_X(t) = \mathrm{E}(X e^{tX})$: $M'_X(0) = \mathrm{E}(X)$.

$M''_X(t) = \mathrm{E}(X^2 e^{tX})$, $M''_X(0) = \mathrm{E}(X^2)$. In general: $M_X^{(n)}(0) = \mathrm{E}(X^n)$.

Although this is basis of name "mgf", it is not often useful in practice: there are easier ways!

(iii) Uniqueness: Mgfs are unique. That is, if $M_X(t) = M_Y(t)$ for all t in an open interval containing 0, then X and Y have the same distribution. **This is useful,** as we will see below.

17.2: Examples of mgf's; Discrete (for convenience, write $z = e^t$).

Binomial: $Bin(n,p)$; $q = 1-p$: $\mathrm{E}(z^X) = \sum_{k=0}^{n} \binom{n}{k} (pz)^k q^{n-k} = (q+pz)^n$

Poisson: $\mathcal{P}o(\mu)$: $\mathrm{E}(z^X) = \sum_{k=0}^{\infty} e^{-\mu}(\mu z)^k/k! = \exp(\mu(z-1))$

Geometric: $Geo(p)$: $\mathrm{E}(z^X) = \sum_{k=1}^{\infty} q^{k-1} p z^k = pz/(1-qz)$

Negative binomial: $NegB(r,p)$:

$$\mathrm{E}(z^X) = \sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r} p^r z^k = (pz)^r \sum_{k=0}^{\infty} \binom{k+r-1}{k} (qz)^k = (pz)^r (1-qz)^{-r}$$

17.3: Examples of mgf's: Continuous

Exponential: $\mathcal{E}(\lambda)$: $\mathrm{E}(e^{tX}) = \lambda \int_0^{\infty} \exp(-(\lambda - t)x)\, dx = \lambda/(\lambda - t)$ provided $t < \lambda$.

Gamma: $G(\alpha, \lambda)$:
$$\mathrm{E}(e^{tX}) = (\lambda^{\alpha}/\Gamma(\alpha)) \int_0^{\infty} x^{\alpha-1} \exp(-(\lambda - t)x)\, dx = (\lambda/(\lambda - t))^{\alpha}$$

Normal: $N(0,1)$: note $-\frac{1}{2}x^2 + tx = -\frac{1}{2}(x-t)^2 + \frac{1}{2}t^2$

$$\mathrm{E}(e^{tX}) = (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} \exp(-\frac{1}{2}x^2 + tx)\, dx = (\exp(\frac{1}{2}t^2)/\sqrt{2\pi}) \int_{-\infty}^{\infty} \exp(-\frac{1}{2}(x-t)^2)\, dx = \exp(\frac{1}{2}t^2)$$

17.4: Mgf of linear functions and sums of independent r.vs

(i) Let $Y = aX + b$: $M_Y(t) = \mathrm{E}(\exp((aX+b)t)) = e^{bt}\mathrm{E}(\exp((at)X) = e^{bt} M_X(at)$.

(ii) Let $X \sim N(\mu, \sigma^2)$, so $X = \sigma Z + \mu$ where $Z \sim N(0,1)$, so

$M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

(iii) Let X and Y be independent random variables: $W = X + Y$:

$M_W(t) = \mathrm{E}(\exp((X+Y)t) = \mathrm{E}(\exp(Xt)\exp(Yt)) = \mathrm{E}(e^{Xt})\mathrm{E}(e^{Yt}) = M_X(t) M_Y(t)$

(iv) Let $X_1, ..., X_n$ be i.i.d. with same dsn as X. $W = \sum_{i=1}^{n} X_i$

$M_W(t) = \prod_{i=1}^{n} M_{X_i}(t) = (M_X(t))^n$.

17.5 Immediate conclusions!!

Sum of independent Binomials (same p) is Binomial;

Sum of independent Poisson (any means) is Poisson

Sum of independent Geometrics (same p) is Negative Binomial; and of NegBin is also NegBin.

Sum of independent Exponentials (same rate) is Gamma; and of Gamma is also Gamma.

Sum of independent Normals (any mean/variance) is Normal; hence any linear combination also Normal.