Asynchronous Lecture

Expectations of Sums

Recall that:

- In the discrete case, $E[X] = \sum_i x_i p(x_i)$,
- In the continuous case, $E[X] = \int x f(x) dx$

If random variables $X, Y$ have joint pdf $f(x, y)$, then

$$E[g(X, Y)] = \int \int g(x, y) f(x, y) dx dy$$

$g(X, Y)$ may be anything from $X$, to $Y$, to $2XY^2$.

Example: $g(x, y) = x + y$

$$E[g(X, Y)] = \int \int (x + y) f(x, y) dx dy$$
$$= \int \int x f(x, y) dx dy + \int \int y f(x, y) dx dy$$
$$= \int x f_x(x) dx + \int y f_y(y) dy$$
$$= E[X] + E[Y]$$

Linearity of expectation $\Rightarrow E[X_1+X_2+\cdots+X_n] = E[X_1]+E[X_2]+\cdots+E[X_n]$
Note: this is not necessarily the case for infinite sums. Only if the following exchange of $\lim_{n \to \infty}$ and $\sum_{i=1}^{n} x_i$ holds is it true.

$$E[\sum_{i=1}^{\infty} x_i] = E[\lim_{n \to \infty} \sum_{i=1}^{n} x_i]$$

$$= \lim_{n \to \infty} \sum_{i=1}^{n} x_i E[x_i]$$

Guaranteed to be exchangeable if $\forall i, x_i \geq 0$; or if $\sum_{i=1}^{n} E[|x_i|] < 0$ (absolutely convergent)

Covariance and Variance of Sums

Definition: Covariance

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

$$= E[XY] - E[X]E[Y]$$

$$= \int \int xy f(x,y)dxdy - E[X]E[Y]$$

If $X$ and $Y$ are independent, then $Cov(X, Y) = 0$

Note: this is not necessarily true in reverse!

Covariance Properties

1. $Cov(X, Y) = Cov(Y, X)$
2. $Cov(X, X) = Var(X)$
3. $Cov(aX, Y) = aCov(X, Y)$
4. $Cov(\sum_{i=1}^{n} x_i, \sum_{j=1}^{m} y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(x_i, y_j)$

If we combine properties 4 and 2,

$$Var(\sum_{i=1}^{n} x_i) = Cov(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(x_i, x_j)$$
We can pull out where \( x_i = x_j \), which occurs when \( i = j \)

\[
= \sum_{i=1}^{n} Var(x_i) + \sum_{i \neq j} Cov(x_i, x_j)
\]

And then add property 1,

\[
= \sum_{i=1}^{n} Var(x_i) + 2 \sum_{i<j} Cov(x_i, x_j)
\]

If the \( x_i \) are independent, \( Var(\sum_{i=1}^{n} x_i) = \sum_{i=1}^{n} Var(x_i) \)

**Correlation**

**Definition:** Correlation

\[
\rho(x, y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}
\]

We can compare this to linear algebra. Recall that the dot (inner) product \( \vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i y_i \), and that the norm \( |\vec{x}| = \vec{x} \cdot \vec{x} \). We can also rewrite the dot product as

\[
\vec{x} \cdot \vec{y} = |\vec{x}| |\vec{y}| \cos \theta
\]

and replace each term with parts from the definition of covariance, such that

\[
\begin{align*}
\vec{x} \cdot \vec{y} &\rightarrow Cov(X, Y) \\
|\vec{x}|^2 &\rightarrow var(x) \\
|\vec{x}| \rightarrow std(x) \\
\cos \theta &\rightarrow \rho(x, y)
\end{align*}
\]

Then, we end up with

\[
\cos \theta = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|}
\]
If \( \rho(x, y) = 0 \), we say that \( x \) and \( y \) are uncorrelated, or in vector terms, they are orthogonal (\( \cos \theta = 0 \Rightarrow \theta = \frac{\pi}{2} \)).

If \( \rho(x, y) = 1 \), we know that \( x = aX + b \), or in vector terms, they are colinear (\( \cos \theta = 1 \Rightarrow \theta = 0 \)).

**Synchronous Lecture**

**Expectations of Sums**

Recall that for a collection of Random Variables \( X_1 + X_2 + \ldots + X_n \), if \( Y = X_1 + X_2 + \ldots + X_n \), then \( E[X] = E[X_1 + X_2 + \ldots + X_n] = \sum E[X_i] \)

**Definition: Sample mean**

\[
\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)
\]

If \( X_i \) are IID Random Variables, with each \( E[x_i] = \mu_i \),

\[
E[\bar{x}] = \frac{1}{n}(\sum E[X_i])
\]

\[
= \frac{1}{n} \cdot \mu
\]

\[
= \mu
\]

note: in statistics, by observing data, we infer property of the R.V. process. For here, we estimate the mean of those Random Variables by taking average of data.

\[
E[g(X, Y)] = \int \int (x + y)f(x, y)dxdy
\]

\[
= \int \int xf(x, y)dxdy + \int \int yf(x, y)dxdy
\]

\[
= \int xf_x(x)dx + \int yf_y(y)dy
\]

\[
= E[X] + E[Y]
\]
Covariance

Definition: *Covariance*

\[ \text{Cov}(X,Y) = E[(X - \mu_x)(Y - \mu_y)] \]

If \( X, Y \) independent, then \( \text{Cov}(X,Y) = 0 \)

Correlation

Definition: *Correlation*

\[ \rho(x,y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} \]

with \( \rho(x,y) \in [-1,1] \).

If we consider \( \text{Cov}(aX,Y) \),

\[ \text{Cov}(aX,Y) = a\text{Cov}(X,Y) \]
\[ \rho(aX,Y) = \frac{a\text{Cov}(X,Y)}{\sqrt{\text{Var}(aX)}\sqrt{\text{Var}(Y)}} \]
\[ = \frac{a\text{Cov}(X,Y)}{\sqrt{a^2\text{Var}(X)}\sqrt{\text{Var}(Y)}} \]
\[ = \frac{a}{|a|}\rho(X,Y) \]

note: correlation shows that we scale the random variables by some characteristic size, and then we compute Covariances.

Sample Variance

Definition: *Sample Variance*

If \( X_1 + X_2 + \cdots + X_n \) are IID Random Variables with mean \( \mu \) and variance \( \sigma^2 \), then the sample variance is given by

\[ S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \]

Note that \( E[S^2] = \sigma^2 \).

Student question about how to visualize these computations. There are three graphs of visualization of different covariances.
Exercises

Theoretical exercise 7.23
If \( Y = a + bX \), what is \( \rho(X, Y) \)?

\[
\begin{align*}
Var(X) &= \sigma^2 \\
Var(Y) &= Var(a + bX) \\
&= b^2 \sigma^2
\end{align*}
\]

\[
Cov(X, Y) = Cov(X, a + bX)
= E(X - \mu)[a + bX - (a + b\mu)]
= E[(X - \mu)(bX - b\mu)]
= bE[(X - \mu)(X - \mu)]
= b\sigma^2
\]

\[
\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}
= \frac{b\sigma^2}{\sigma \cdot |b|\sigma}
= \frac{b}{|b|}
\]

1 if \( b > 0 \), and -1 if \( b < 0 \)

Theoretical exercise 7.4
Let \( X \) be a Random Variable with

\[
\begin{align*}
E[X] &= \mu < \infty \\
Var[X] &= \sigma^2 < \infty
\end{align*}
\]
and $g = g(x)$ is twice differentiable, how to approximate $E[g(x)]$?

By Taylor Series, we know that

\[
g(x) \approx g(\mu) + g'(\mu)(x - \mu) + \frac{g''(\mu)}{2}(x - \mu)^2\]

\[
E[g(x)] \approx E[g(\mu) + g'(\mu)(x - \mu) + \frac{g''(\mu)}{2}(x - \mu)^2]
\]

\[
= g(\mu) + g'(\mu)E[X - \mu] + \frac{g''(\mu)}{2}E[(X - \mu)^2]
\]

\[
= g(\mu) + \frac{g''(\mu)}{2}\sigma^2
\]

\[
E[g(x)] \approx g(\mu) + \frac{g''(\mu)}{2}\sigma^2
\]