Today: Covariance

Next: ASV 9.1

Week 10:

- Homework 7 due Friday December 2
- CAPES
- Final exam: Tuesday, December 6, 3 - 6 PM
Corollary If $X_1, \ldots, X_n$ are independent, then

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$$

Proof $(n=2)$. Suppose that $X, Y$ are independent, $E(X) = \mu$, $E(Y) = \lambda$. Then

$$\text{Var}(X+Y) = E\left[ (X+Y - E[X+Y])^2 \right]$$
Variance of a sum of independent random variables

Example

Binomial: \(X_1, ..., X_n\) independent identically distributed (iid)

\[X_i \sim Ber(p), \ Var(X_i) = p(1-p), \ Sn = X_1 + \cdots + X_n\]

\[\text{Var}(S_n) = \]

Sample mean: \(X_1, ..., X_n\) independent identically distributed (iid)

\[\mathbb{E}(X_i) = \mu, \ Var(X_i) = \sigma^2\]

\[\mathbb{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) = , \ Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \]
**Coupon collector's problem**

There is a toy in each box of cereal. There are $n$ different toys, distributed equally likely and independently among the boxes. Let $T_n$ denote the number of boxes you need to buy to collect all $n$ toys. Compute $E(T_n)$, $\text{Var}(T_n)$.

**Option 1**: determine the distribution of $T_n$ → difficult/unnecessary

**Option 2**: decompose $T_n$ into simpler random variables

- denote by $T_k$ the number of boxes you buy before you collect $k$ different toys ($T_1 = 1$)
- $T_k$'s are not independent. What about $W_k := T_{k+1} - T_k$?
- $W_k := T_{k+1} - T_k = \# \text{ boxes you buy to get a new toy}$
- What is the distribution of $W_k$? $W_k = T_{k+1} - T_k$
Coupon collector's problem

- \( W_k = T_{k+1} - T_k \sim \text{Geom} \left( \frac{n-k}{n} \right) \)
- \( T_n = T_1 + (T_2 - T_1) + (T_3 - T_2) + \ldots + (T_n - T_{n-1}) \)

- \( E(W_k) = \frac{n}{n-k} \), \( E(T_n) \)

- If \( X \sim \text{Geom}(p) \), \( \text{Var}(X) = \frac{1-p}{p^2} \), so

- \( \text{Var}(W_k) = \)
- \( \text{Var}(T_n) = \)
Def. (Convolution of distributions - Section 7)

Let $X$ and $Y$ be random variables. Then the distribution of $X+Y$ is called the **convolution** of the distributions of $X$ and $Y$.

If $X$ and $Y$ are continuous and $f_X$ and $f_Y$ are their PDFs, then the PDF of $X+Y$ is given by

$$f_{X+Y}(s) = f_X * f_Y(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s-x) \, dx = \int_{-\infty}^{\infty} f_X(s-y) f_Y(y) \, dy$$

(similar formula for discrete random variables)

If $X$ and $Y$ are independent, it may be easier to compute the
Moment generating function of a sum of indep. RVs

Let $X, Y$ be independent random variables. Then the MGF of $X+Y$ is

1) $X \sim \text{Poisson}(\lambda), Y \sim \text{Poisson}(\mu)$, independent. Distribution of $X+Y$?

2) $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, independent. Distribution of $X+Y$?
Covariance

Suppose that we have a random variable $X$.

- $E(X)$ - mean value, average of a large number of independent realizations
- $\text{Var}(X)$ - variance, fluctuations of $X$, how far the realizations are spread around the mean

Covariance describes strength and type of dependence between two random variables.

**Def.** Let $X$ and $Y$ be random variables defined on the same probability space. The covariance of $X$ and $Y$ is given by

**Computation:**
Covariance

Example  Let $X, Y$ be discrete random variables with the joint PMF $P(X=k, Y=\ell)$ given by the table

<table>
<thead>
<tr>
<th>$k$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>0</td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.4</th>
<th>0.4</th>
<th>0.2</th>
</tr>
</thead>
</table>

$E(X) =$  
$E(Y) =$  
$E(XY) =$  
$\text{Cov}(X,Y) =$
Some heuristics

By definition, $\text{Cov}(X,Y) = \mathbb{E}[(X-E(X))(Y-E(Y))]$

- $(X-E(X))(Y-E(Y))$ is positive if $(X-E(X))$ and $(Y-E(Y))$
- $(X-E(X))(Y-E(Y))$ is negative if $(X-E(X))$ and $(Y-E(Y))$

Thus,

- $\text{Cov}(X,Y) > 0$ means that on average $X-E(X)$ and $Y-E(Y)$ have
- $\text{Cov}(X,Y) < 0$ means that on average $X-E(X)$ and $Y-E(Y)$ have
- If $\text{Cov}(X,Y) = 0$, we say that
Example

Let $(X, Y)$ be a uniformly distributed random point on the trapezoid with vertices $(0,0), (2,0), (1,1), (0,1)$.

Is $\text{Cov}(X, Y)$

(a) positive?

(b) negative?

Joint density:

$E(X) =$

$E(Y) =$

$E(XY) =$

$\text{Cov}(X, Y) =$
Variance of a sum

Thm. Let $X_1, \ldots, X_n$ be random variables with finite variances.

Then $\text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i)$.

For example, $\text{Var}(X+Y) = \ldots$

Example. Urn has 5 balls: 3 red and 2 green. Draw 2 balls and let $X =$ number of red ball. Compute $\text{Var}(X)$ if the sampling is done (a) with replacement (b) without replacement.

Let $Y_1 = \begin{cases} 1, & \text{first ball is red} \\ 0, & \text{otherwise} \end{cases}$, $Y_2 = \begin{cases} 1, & \text{second ball is red} \\ 0, & \text{otherwise} \end{cases}$.

$Y_1 \sim \text{Ber}(\frac{3}{5})$, $Y_2 \sim \text{Ber}(\frac{3}{5})$, $\text{Var}(Y_1) = \text{Var}(Y_2) = \frac{6}{25}$.

(a) $\text{Var}(X) = \ldots$

(b) $E(Y_1 Y_2) = \ldots$
Uncorrelated vs Independent

- $X$ and $Y$ are independent $\Rightarrow \text{Cov}(X, Y) = 0$
  $$E(XY)$$

- $\text{Cov}(X, Y) = 0 \n\not\Rightarrow X$ and $Y$ are independent

Example of random variables $X, Y$ that are not independent, but $\text{Cov}(X, Y) = 0$