Week 7
Oct. 15 – Oct. 19
Important Discrete Distributions

Lecture 17. Bernoulli and Binomial, Expectation
Inclusion and Exclusion Principle:

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= \sum_{i=1}^{n} P(A_i) \\ &\quad - \sum_{1\le i<j\le n}^{n} P(A_i \cap A_j) \\ &\quad + \sum_{1\le i<j<k\le n}^{n} P(A_i \cap A_j \cap A_j) - \cdots \\ &\quad + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n) \end{aligned}$$

Hat Check problem:
a hat-check girl in a restaurant, having checked n hats, gets them hopelessly scrambled and returns them at random to the n owners as they leave. What is the probability that nobody gets his own hat back?

Hint: Define

$$A_i = \{i\text{th owner gets his own hat}\}$$

Then

$$\begin{aligned} P(A_i) &= \frac{(n-1)!}{n!} = \frac{1}{n} \\ P(A_i \cap A_j) &= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)},\ i<j \\ P(A_i \cap A_j \cap A_k) &= \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)},\ i<j<k \\ &\cdots \end{aligned}$$

and

$$\begin{aligned} &P(A_1 \cup A_2 \cup \cdots \cup A_n) \\ =\ & 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!}. \end{aligned}$$

Review

Probability. Conditional probability. Combinatorics.

Sample space. Event. Probability. Random variable. Expectation. Conditional distribution (density). $\binom{n}{k}$.

Clarification: Let A be an event. Define

$$\{A\} = \begin{cases} 1 & \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

This is a function of outcome, so it is a random variable. An alternative notation for $\{A\}$ is $1_A(\omega)$. Note that

$$\begin{aligned} P(\{A\} = 1) &= P(\omega \in A) = P(A) \\ P(\{A\} = 0) &= P(\omega \notin A) = 1 - P(A) \end{aligned}$$

then the expectation of $\{A\}$ is

$$\mathbb{E}\{A\} = 1 \cdot P(A) + 0 \cdot (1 - P(A)) = P(A).$$

What to do this week: we describe the discrete probability distributions and continuous probability distributions that occur most often in the analysis of experiments.

Discrete uniform distribution.

We have seen some examples that all outcomes of an experiment are equally likely. Let X be a random variable representing the outcome of an experiment of this kind. Let $x_1, x_2, \ldots, x_m$ denote all possible outcomes. Then

$$\Omega = \{x_1, x_2, \ldots, x_m\}$$

and

$$P(X = x_i) = \frac{1}{m}.$$

A simple but important example

$$\Omega = \{0, 1\}$$

and

$$P(X = 0) = P(X = 1) = 1/2.$$

Binomial.

It is the distribution of random variable which counts the number of heads when a coin is tossed n times. Let $X_i = 1$ if a head occurs at the ith toss, otherwise $X_i = 0$. Assume that

$$P(X_i = 1) = p,\ P(X_i = 0) = 1 - p,$$

then the distribution of $X = X_1 + X_2 + \ldots + X_n$ is

$$P(X = x) = \binom{n}{x} p^x q^{n-x},\ x = 0, 1, \ldots, n.$$

Question:

$$\mathbb{E}X = ?$$

Solution 1:

$$\begin{aligned}
\mathbb{E}X &= \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} x \cdot \frac{n!}{x!\,(n-x)!} p^x q^{n-x} \\
&= \sum_{x=1}^{n} np \cdot \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1} q^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1} q^{n-x} = ?
\end{aligned}$$

Solution 2:

$$\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + \ldots + X_n) \stackrel{?}{=} \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_n = ?$$

For solution 2 we need to answer a question: for two discrete (continuous) random variables X and Y, is it true that

$$\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y?$$

Proof: Let the sample space of X and Y be denoted by Ω_X and Ω_Y, and

$$\begin{aligned}
\Omega_X &= \{x_1, x_2, \ldots, x_i, \ldots\} \\
\Omega_Y &= \{y_1, y_2, \ldots, y_j, \ldots\}.
\end{aligned}$$

The sample space of (X, Y) is

$$\Omega_{(X,Y)} = \{(x_i, y_j), i = 1, 2, \ldots,\ j = 1, 2, \ldots\}.$$

Then

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_i \sum_j (x_i + y_j) P(X = x_i, Y_j = y_j) \\
&= \sum_i \sum_j x_i P(X = x_i, Y_j = y_j) + \sum_i \sum_j y_j P(X = x_i, Y_j = y_j) \\
&= \sum_i x_i \sum_j P(X = x_i, Y_j = y_j) + \sum_j y_j \sum_i P(X = x_i, Y_j = y_j) \\
&= \sum_i x_i P(X = x_i) + \sum_j y_j P(Y_j = y_j) \\
&= \mathbb{E}X + \mathbb{E}Y.
\end{aligned}$$

Question:

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_n$$

Question:

$$\mathbb{E}cX = c\mathbb{E}X?$$

Question:

$$\mathbb{E}\left(X - \mathbb{E}X\right)^2 = \mathbb{E}X^2 - \left(\mathbb{E}X\right)^2?$$

Question: for two continuous random variables X and Y, is it true that

$$\mathbb{E}\left(X + Y\right) = \mathbb{E}X + \mathbb{E}Y?$$

The answer is exactly the same:

$$\mathbb{E}\left(X + Y\right) = \int\int\left(x + y\right) f_{X,Y}\left(x, y\right) dxdy$$

where $f_{X,Y}\left(x, y\right)$ is the joint density of (X, Y). Then

$$\begin{aligned}
\mathbb{E}\left(X + Y\right) &= \int\int x f_{X,Y}\left(x, y\right) dxdy + \int\int y f_{X,Y}\left(x, y\right) dxdy \\
&= \int x\left[\int f_{X,Y}\left(x, y\right) dy\right] dx + \int y\left[\int f_{X,Y}\left(x, y\right) dy\right] dx \\
&= \int x f_X\left(x\right) dx + \int y f_Y\left(y\right) dx \\
&= \mathbb{E}X + \mathbb{E}Y.
\end{aligned}$$

Lecture 18. Bernoulli and Binomial. Variance
Binomial.

$$P\left(X=x\right)=\binom{n}{x}p^{x}q^{n-x},\ x=0,1,\ldots,n.$$

Question:

$$\mathbb{E}\left(X-\mathbb{E}X\right)^{2}=?$$

or

$$\mathbb{E}X^{2}=?$$

Variance

$$var\left(X\right)=\mathbb{E}\left(X-\mathbb{E}X\right)^{2}.$$

Do you like to calculate

$$\begin{aligned}&\sum_{x=0}^{n}\left(x-np\right)^{2}\cdot\frac{n!}{x!\left(n-x\right)!}p^{x}q^{n-x}\\ =&\sum_{x=1}^{n}\left(x-np\right)^{2}\cdot\frac{n!}{x!\left(n-x\right)!}p^{x}q^{n-x}?\end{aligned}$$

For $n=1$, it is easy to see

$$\mathbb{E}\left(X-p\right)^{2}=\left(1-p\right)^{2}p+p^{2}\left(1-p\right)=p\left(1-p\right).$$

Similar trick,

$$\begin{aligned}&\mathbb{E}\left(X_{1}-p+X_{2}-p+\ldots+X_{n}-p\right)^{2}\\ =&\mathbb{E}\left(X_{1}-p\right)^{2}+\mathbb{E}\left(X_{2}-p\right)^{2}+\ldots+\mathbb{E}\left(X_{n}-p\right)^{2}?\end{aligned}$$

The right hand side is

$$\begin{aligned}&\mathbb{E}\left(X_{1}-p\right)^{2}+\mathbb{E}\left(X_{2}-p\right)^{2}+\ldots+\mathbb{E}\left(X_{n}-p\right)^{2}\\ &+\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{E}\left(X_{i}-p\right)\left(X_{j}-p\right)\end{aligned}$$

Is it true

$$\mathbb{E}\left(X_{i}-p\right)\left(X_{j}-p\right)=\mathbb{E}\left(X_{i}-p\right)\cdot\mathbb{E}\left(X_{j}-p\right)=0?$$

Yes. It is true. More generally, for two independent discrete (continuous) random variables X and Y, is it true that

$$\mathbb{E}g\left(X\right)h\left(Y\right)=\mathbb{E}g\left(X\right)\cdot\mathbb{E}h\left(Y\right).$$

Proof:

$$\begin{aligned}\mathbb{E}g(X)h(Y) &= \int\int g(x)h(y)f(x,y)\,dxdy \\ &= \int\int g(x)h(y)f_X(x)f_Y(y)\,dxdy \\ &= \int g(x)f_X(x)\,dx\int h(y)f_Y(y)\,dxdy \\ &= \mathbb{E}g(X)\cdot\mathbb{E}h(Y).\end{aligned}$$

Theorem Let X_1, $X_2, \ldots, X_n$ be n independent and identically distributed random variables, and $\mathbb{E}X_i = \mu$ and $\mathbb{E}(X_i-\mu)^2 = \sigma^2$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\begin{aligned}\mathbb{E}S_n &= n\mu \\ var(S_n) &= \mathbb{E}(S_n - n\mu)^2 = n\sigma^2.\end{aligned}$$

Question: Let

$$S_n^* = \frac{\mathbb{E}_n - n\mu}{\sqrt{n}\sigma}$$

Then

$$\begin{aligned}\mathbb{E}S_n^* &= ? \\ var(S_n^*) &= ?\end{aligned}$$

Lecture 19. Geometric and Negative Binomial. Expectation and Variance.

Geometric.

Flip a coin until you see a head. Let T_1 be the number of flips until you see the head.

$$P(T_1 = x) = q^{x-1}p,\ x \geq 1$$

Question:

$$\begin{aligned} \mathbb{E}X &= ? \\ \mathbb{E}(X - PX)^2 &= ? \end{aligned}$$

By definition

$$\begin{aligned} \mathbb{E}X &= \sum_{x=1}^{\infty} x q^{x-1} p \\ &= p\frac{d}{dq}\left(\sum_{x=0}^{\infty} q^x\right) \\ &= p\frac{d}{dq}\frac{1}{1-q} = 1/p. \end{aligned}$$

Homework problem

$$Var(X) = \frac{1-p}{p^2}.$$

Negative Binomial.

Flip a coin until you see k heads. Let T be the number of flips until you see k heads.

$$P(T = x) = \binom{x-1}{k-1} p^k q^{x-k},\ x \geq k$$

Question:

$$\begin{aligned} \mathbb{E}X &= ? \\ \mathbb{E}(X - PX)^2 &= ? \end{aligned}$$

Note that

$$T = T_1 + T_2 + \ldots + T_k.$$

Hypergeometric.

Suppose that we have a set of n balls, of which k are red and $N - k$ are blue. We choose n balls, without replacement, and define X to be the number of the red balls in our sample.

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Example 5.6 in the textbook.

Observed tables

	Democrat	Republican	
Female	24	4	28
Male	8	14	22
	32	18	50

Expected table

	Democrat (*yellow*)	Republican(*red*)	
Female	18	10	28
Male	14	8	22
	32	18	50

If we choose 28 balls out of 50, we should expect to see, on the average, the same percentage of yellow balls in our sample as in the urn, if there is no association between gender and political party. Thus we see expect to see $28\frac{32}{50} \approx 17.92 \approx 18$ yellow balls in our sample. But the actual number in our data is 24. What is the probability to have an observation ≥ 24? It is just .000395.