STAT 230

July 4 Tutorial Solutions

Preview

In this tutorial, we will look at discrete distributions, their pf's and cdf's and how to use them to calculate probabilities. A general application for a discrete r.v. will be explored and we will also explore expectation and variance for a discrete r.v.

We will also look at continuous distributions, their pf's and cdf's and how to use them to calculate probabilities. We will explore percentiles, and variance for a continuous r.v.

Problem 1

Suppose you are given the following information regarding the pf of a discrete random variable, X:

x	0	1	2	3	4	5	6
f(x)	0.05	0.1	0.1	0.2	0.3	0.05	0.2

a) Using the information given above, determine $P(2 \leq X < 5)$.

b) Determine the cdf of the discrete random variable X.

c) Calculate the expected value and variance of X.

d) Suppose $Y = 2 - 3X$. Calculate the expected value and variance of Y.

Problem 1 Solution

Suppose you are given the following information regarding the pf of a discrete random variable, X:

x	0	1	2	3	4	5	6
f(x)	0.05	0.1	0.1	0.2	0.3	0.05	0.2

a) Using the information given above, determine $P(2 \leq X < 5)$.

Solution:

$P(2 \leq X < 5) = P(X = 2) + P(X = 3) + P(X = 4) = 0.1 + 0.2 + 0.3 = 0.6$

Problem 1 Solution (cont'd)

Suppose you are given the following information regarding the pf of a discrete random variable, X:

x	0	1	2	3	4	5	6
f(x)	0.05	0.1	0.1	0.2	0.3	0.05	0.2

b) Determine the cdf of the discrete random variable X.

Solution:

Using the pf table, we can easily obtain the corresponding cdf table. Remember that the cdf is simply accumulating probability as we move along the x-axis (values of X).

To start, with f(0) = 0.05, we have F(0) = 0.05. Next f(x) = 0.1, so F(1) = F(0) + f(1) = 0.15.

Then, F(2) = F(1) + f(2) = 0.15 + 0.1 = 0.25. F(3) = F(2) + f(3) = 0.25 + 0.2 = 0.45.

F(4) = F(3) + f(4) = 0.45 + 0.3 = 0.75. F(5) = F(4) + f(5) = 0.75 + 0.05 = 0.8,

and finally F(6) = F(5) + f(6) = 0.8 + 0.2 = 1, as expected.

Problem 1 Solution (cont'd)

Here is the resulting cdf (in table form)

x	0	1	2	3	4	5	6
F(x)	0.05	0.15	0.25	0.45	0.75	0.8	1

Problem 1 Solution (Cont'd)

Suppose you are given the following information regarding the pf of a discrete random variable, X:

x	0	1	2	3	4	5	6
f(x)	0.05	0.1	0.1	0.2	0.3	0.05	0.2

c) Calculate the expected value and variance of X.

Solution:

E(X) = $\sum_{x=0}^{6} x * f(x) = 0(0.05) + 1(0.1) + 2(0.1) + 3(0.2) + 4(0.3) + 5(0.05) + 6(0.2) = 3.55$

Var(X) = $E[X^2] - (E[X])^2$

$E[X^2]$ **=** $\sum_{x=0}^{6} x^2 * f(x) = (0)^2(0.05) + (1)^2(0.1) + (2)^2(0.1) + (3)^2(0.2) + (4)^2(0.3) + (5)^2(0.05) + (6)^2(0.2)$ **= 15.55**

So, Var(X) = 15.55 - $(3.55)^2$ **= 2.9475**

Problem 1 Solution (Cont'd)

d) Suppose Y = 2 – 3X. Calculate the expected value and variance of Y.

Solution:

E[Y] = E[2 - 3X] = 2 – 3E[X] = 2 – 3(3.55) = -8.65

Note: The expected value of a constant is simply the value of the constant.

Var[Y] = Var[2 – 3X] = $(-3)^2$Var(X) = 9*Var(X) = 9*(2.9475) = 26.5275

Note: Make sure that you square the coefficient in front of X when determining the Variance. Remember that Variance cannot be negative! Also, remember that the variance of a constant is 0. Constants don't vary.

Problem 2

In a game of “green ball”, two selections are made, without replacement from a box that has 4 white ping pong balls and 2 green ping pong balls. The amount that a player wins is determined by how many green ping pong balls are selected. A player will pay $5 to play. They can decide to be paid under two rules. The information is given below:

Rule A

# of green balls	Amount Received
0	0
1	$2
2	$10

Rule B

# of green balls	Amount Received
0	0
1	$1
2	$20

Is this game fair?

If you were playing the game, which rule would you choose, and why?

Problem 2 Solution

In a game of “green ball”, two selections are made, without replacement from a box that has 4 white ping pong bolls and 2 green ping pong balls. The amount that a player wins is determined by how many green ping pong balls are selected. A player will pay $5 to play. They can decide to be paid under two rules.

Is this game fair?

Solution

In this case, we can let the net winnings, N = W – 5, where W = amount won. Remember that it costs $5 to play also.

We will look at E(N) = E(W) – 5 to answer these questions.

Rule A: E(N) = $(0)*\frac{\binom{2}{0}\binom{4}{2}}{\binom{6}{2}} + (2)*\frac{\binom{2}{1}\binom{4}{1}}{\binom{6}{2}} + (10)*\frac{\binom{2}{2}\binom{4}{0}}{\binom{6}{2}} - 5 = -49/15$

Rule B: E(N) = $(0)*\frac{\binom{2}{0}\binom{4}{2}}{\binom{6}{2}} + (1)*\frac{\binom{2}{1}\binom{4}{1}}{\binom{6}{2}} + (20)*\frac{\binom{2}{2}\binom{4}{0}}{\binom{6}{2}} - 5 = -47/15$

For the game to be considered fair, the expected net winnings / expected gain would need to be 0. So neither game is considered to be fair.

If you were playing the game, which rule would you choose, and why?

As noted above, you really shouldn’t play this game at all! However, if you had to choose, Rule B should be chosen as the expected net winnings are slightly better.

Problem 3

Suppose that the length of time X (in years) that a patient, suffering from a certain disease, is in remission has a pdf given by: $f_X(x) = \frac{x^2}{9}$ for $0 \leq x \leq 3$; 0, otherwise.

a) Determine the cdf of X.
b) Use the cdf from part a) to determine $P(X > 2)$.
c) Calculate Var(X).
d) What is the median length of time (in years) that a patient, suffering from a certain disease, is in remission?

Problem 3 Solution

Suppose that the length of time X (in years) that a patient suffering from a certain disease is in remission has a pdf given by:

$$f_X(x) = \frac{x^2}{9} \text{ for } 0 \le x \le 3; \; 0, \text{ otherwise.}$$

a) Determine the cdf of X.

Solution:

The cdf of X, or F(x) = P(X ≤ x) = $\int_0^x f(s)ds = \int_0^x \frac{s^2}{9} ds = \left[\frac{s^3}{27}\right] |_0^x$

So, F(x) = 0 for x ≤ 0, F(x) = $\frac{x^3}{27}$ **for 0 < x ≤ 3, and F(x) = 1 for x > 3.**

Notes: Make sure that you put a "dummy" variable in the integration. Here, *s* was chosen, but you could choose any letter other than *x* in this case. Also, ensure that you don't simply integrate the pdf from 0 to 3. You will just get 1, verifying that this pdf integrates to 1, as it should, if it's valid.

Problem 3 Solution (cont'd)

b) Use the cdf from part a) to determine P(X > 2).

Solution:

From part a), F(x) = $\frac{x^3}{27}$ for 0 < x ≤ 3.

We can use this portion of the definition of the cdf to determine the required probability, or P(X > 2).

In this case, P(X > 2) = 1 – P(X ≤ 2) = 1 - F(2) = 1 - $\frac{2^3}{27}$ = $\frac{19}{27}$ = 0.704

Problem 3 Solution (cont'd)

c) Calculate Var(X).

Solution:

Var(X) = $\boldsymbol{E\left[X^2\right] - (E[X])^2}$

E(X) = $\boldsymbol{\int_0^3 x * f(x)dx = \int_0^3 \frac{x^3}{9} dx = \left[\frac{x^4}{36}\right] |_0^3}$ **=** $\boldsymbol{\frac{81}{36} = \frac{9}{4}}$

Similarly, E($\boldsymbol{X^2}$**) =** $\boldsymbol{\int_0^3 x^2 * f(x)dx = \int_0^3 \frac{x^4}{9} dx = \left[\frac{x^5}{45}\right] |_0^3}$ **=** $\boldsymbol{\frac{243}{45} = \frac{27}{5}}$

So, Var(X) = $\boldsymbol{\frac{27}{5}}$ **-** $\boldsymbol{\left(\frac{9}{4}\right)^2}$ **=** $\boldsymbol{\frac{27}{80}}$

Problem 3 Solution (cont'd)

d) What is the median length of time (in years) that a patient, suffering from a certain disease, is in remission?

Solution:

We want the median of the distribution of X in this case. To find the median (or 50th percentile) of the distribution of X, we simply need to set the cdf equal to 0.5 and solve for the median value, x.

From part a), the cdf, F(x) = P(X ≤ x) = F(x) = $\frac{x^3}{27}$ **for 0 < x ≤ 3**

So, we have $\frac{x^3}{27}$ **= 0.5 =>** $x^3 = \frac{27}{2}$**. So, x =** $\sqrt[3]{27/2}$ **= 2.381**

Put another way, P(X ≤ $\sqrt[3]{27/2}$ **) = 0.5**