Discrete Random Variables (Ch. 5.1)

Dason Kurkiewicz

Iowa State University

June 5, 2013

Outline

What is a random variable?

Probability

Probability Mass Functions (pmf)

Cumulative Distribution Functions (cdf)

Expected Value

Variance and Standard Deviation

What is a random variable?

- **Random variable**; a quantity that can be thought of as dependent on chance phenomena.
  - $X =$ the value of a coin toss (heads or tails).
  - $Z =$ the amount of torque required to loosen the next bolt.
  - $T =$ the time you'll have to wait for the next bus home.
  - $N =$ the number of defective widgets in manufacturing process in a day.
  - $S =$ the number of provoked shark attacks off the coast of Florida next year.
- Two types:
  - **Discrete random variable**: one that can only take on a set of isolated points (X, N, and S).
  - **Continuous random variable**: one that can fall in an interval of real numbers (T and Z).

Discrete random variables

- A discrete random variable has a list of possible values:
  - X = roll of a 6-sided fair die = 1, 2, 3, 4, 5, or 6.
  - Y = roll of a 6-sided *unfair* die = 1, 2, 3, 4, 5, or 6.
- But how do you distinguish between X and Y?

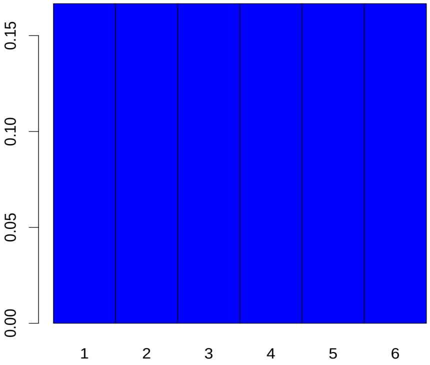
Outline

Probability

- $P(X = x)$, the **probability** that X equals x, is the fraction of times that X will land on x
  1. We expect a fair die to land the number 3 roughly one out of every 6 tosses. Thus, $P(X = 3) = 1/6$
  2. Suppose the unfair die is weighted so that the number 3 only lands one out of every 22 tosses. Then, $P(Y = 3) = 1/22$.

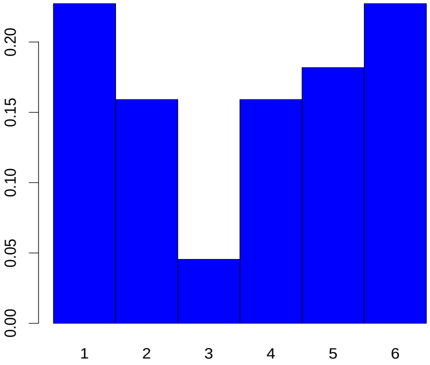
- X has the following probabilities:

x	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6



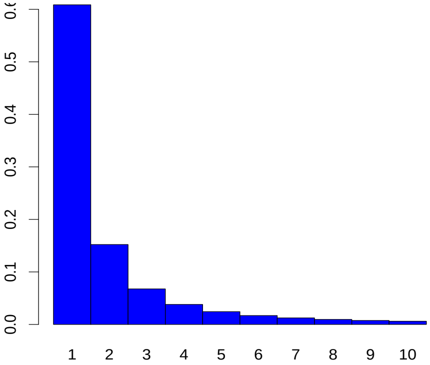
- Say Y has the probabilities:

y	1	2	3	4	5	6
$P(Y=y)$	$5/22$	$7/44$	$1/22$	$7/44$	$2/11$	$5/22$



- S, the number of provoked shark attacks off FL next year, has infinite number of possible values. Here is one possible (made up) distribution:

s	1	2	3	$\cdots$	k	$\cdots$
$P(S = s)$	$\frac{6}{\pi^2}$	$\frac{1}{4}\frac{6}{\pi^2}$	$\frac{1}{9}\frac{6}{\pi^2}$	$\cdots$	$\frac{1}{k^2}\frac{6}{\pi^2}$	$\cdots$



Outline

Probability mass functions (pmf)

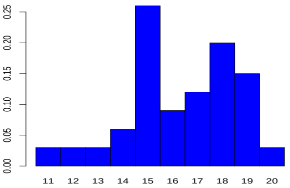
- The **probability mass function (pmf)** $f(x)$ of a random variable X is just $P(X = x)$
  - X has $f(x) = 1/6$
  - S has $f(s) = \frac{1}{2^s}\frac{6}{\pi^2}$.
- We could also write f_X for the pmf of X and f_S for the pmf of S.
- Rules of the pmf f:
  - $f(x) \geq 0$ for all x.
  - $\sum_x f(x) = 1$.

Your turn: calculating probabilities

- Let Z = the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

z	11	12	13	14	15
$f(z) = P(Z = z)$	0.03	0.03	0.03	0.06	0.26

z	16	17	18	19	20
$f(z) = P(Z = z)$	0.09	0.12	0.20	0.15	0.03



- Calculate:
  1. $P(Z \leq 14)$
  2. $P(Z > 16)$
  3. $P(Z \text{ is an even number})$
  4. $P(Z \text{ in } \{15, 16, 18\})$

Answers: calculating probabilities

1.

$$\begin{aligned}
P(Z \leq 14) &= P(Z = 11 \text{ or } Z = 12 \text{ or } Z = 13 \text{ or } Z = 14) \\
&= P(Z = 11) + P(Z = 12) + P(Z = 13) + P(Z = 14) \\
&= f(11) + f(12) + f(13) + f(14) \\
&= 0.03 + 0.03 + 0.03 + 0.06 \\
&= 0.15
\end{aligned}$$

2.

$$\begin{aligned}
P(Z > 16) &= P(Z = 17 \text{ or } Z = 18 \text{ or } Z = 19 \text{ or } Z = 20) \\
&= P(Z = 17) + P(Z = 18) + P(Z = 19) + P(Z = 20) \\
&= f(17) + f(18) + f(19) + f(20) \\
&= 0.12 + 0.20 + 0.15 + 0.03 \\
&= 0.5
\end{aligned}$$

Answers: calculating probabilities

3.

$$\begin{aligned}
P(Z \text{ even}) &= P(Z = 12 \text{ or } Z = 14 \text{ or } Z = 16 \text{ or } Z = 18 \text{ or } Z = 20) \\
&= P(Z = 12) + P(Z = 14) + P(Z = 16) + P(Z = 18) \\
&\quad + P(Z = 20) \\
&= f(12) + f(14) + f(16) + f(18) + f(20) \\
&= 0.03 + 0.06 + 0.09 + 0.20 + 0.03 \\
&= 0.41
\end{aligned}$$

4.

$$\begin{aligned}
P(Z \text{ in } \{15, 16, 18\}) &= P(Z = 15 \text{ or } Z = 16 \text{ or } Z = 18) \\
&= P(Z = 15) + P(Z = 16) + P(Z = 18) \\
&= f(15) + f(16) + f(18) \\
&= 0.26 + 0.09 + 0.02 \\
&= 0.37
\end{aligned}$$

Outline

The cumulative distribution function (cdf)

- **Cumulative distribution function (cdf)**: a function, F, defined by:

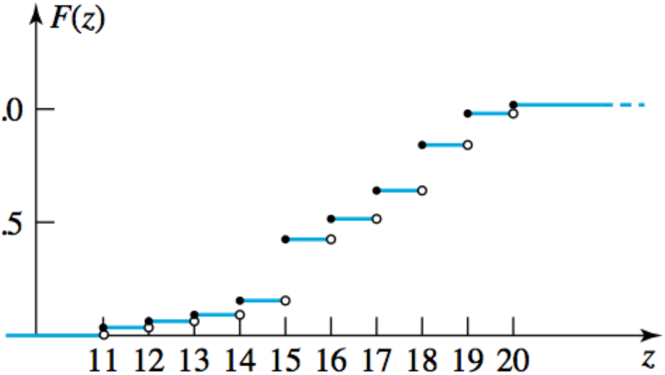
$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \sum_{z \leq x} f(z) \end{aligned}$$

- F has the following properties:
  - $F(x) \geq 0$ for all real numbers x.
  - F is monotonically increasing.
  - $\lim_{x \to -\infty} F(x) = 0$
  - $\lim_{x \to \infty} F(x) = 1$
- When statisticians say "distribution", they mean cdf.

Example: torque random variable, Z

z, Torque	$f(z) = P[Z = z]$	$F(z) = P[Z \leq z]$
11	.03	.03
12	.03	.06
13	.03	.09
14	.06	.15
15	.26	.41
16	.09	.50
17	.12	.62
18	.20	.82
19	.15	.97
20	.03	1.00

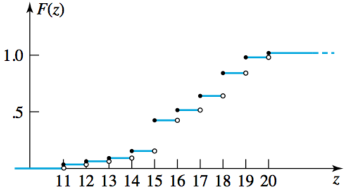
Example: torque random variable, Z



Your turn: calculating probabilities

z	11	12	13	14	15
$F(z) = P(Z \leq z)$	0.03	0.06	0.09	0.15	0.41

z	16	17	18	19	20
$F(z) = P(Z \leq z)$	0.50	0.62	0.82	0.97	1



- Using the cdf only, calculate:
  1. F(10.7)
  2. $P(Z \leq 15.5)$
  3. $P(12.1 < Z \leq 14)$
  4. $P(15 \leq Z < 18)$

Answers: calculating probabilities

1. $F(10.7) = P(Z \leq 10.7) = 0$
2. $P(Z \leq 15.5) = P(Z \leq 15) = 0.41$
3. $$\begin{aligned} P(12.1 < Z \leq 14) &= P(Z = 13 \text{ or } 14) \\ &= f(14) + f(13) \\ &= [f(14) + f(13) + f(12) + f(11)] \\ &\quad - [f(12) + f(11)] \\ &= P(Z \leq 14) - P(Z \leq 12) \\ &= F(14) - F(12) \\ &= 0.15 - 0.06 \\ &= 0.09 \end{aligned}$$

Answers: calculating probabilities

4.

$$
\begin{aligned}
P(15 \le Z < 18) &= P(Z = 15,\ 16,\ \text{or } 17) \\
&= P(Z \le 17) - P(Z \le 14) \\
&= F(17) - F(14) \\
&= 0.62 - 0.15 \\
&= 0.47
\end{aligned}
$$

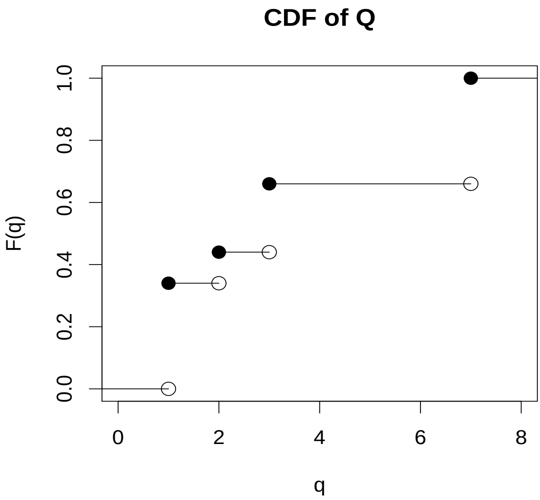
Your turn: drawing the cdf

- Say we have a random variable Q with pmf:

q	1	2	3	7
$f(q)$	0.34	0.1	0.22	0.34

- Draw the cdf.

Answer: drawing the cdf



Outline

Expected Value

- The **expected value** $E(X)$ (also called μ) of a random variable X is given by:

$$\sum_x x \cdot f(x)$$

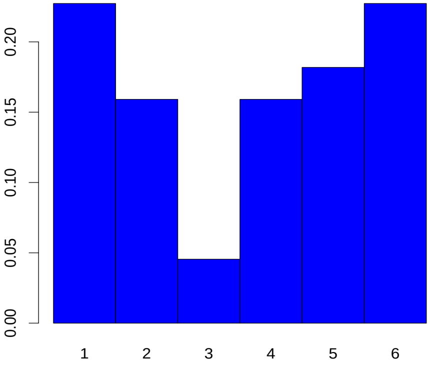
- When X is the roll of a fair die,

$$\begin{aligned} E(X) &= 1f(1) + 2f(2) + 3f(3) + 4f(4) + 5f(5) + 6f(6) \\ &= 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) \\ &= \frac{1+2+3+4+5+6}{6} \\ &= 3.5 \end{aligned}$$

- $E(X)$ is a *weighted average* of the possible values of X, weighted by their probabilities.
- $E(X)$ is the **mean of the distribution** of X

Your turn: expected value

y	1	2	3	4	5	6
$P(Y=y)$	$5/22$	$7/44$	$1/22$	$7/44$	$2/11$	$5/22$

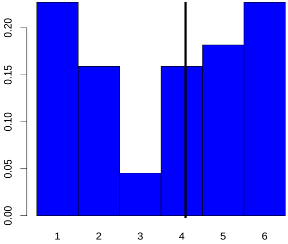


- Calculate $E(Y)$, the expected value of a toss of the *unfair* die.

Answer: expected value

- $$E(Y) = 1(5/22) + 2(7/44) + 3(1/22) \\ + 4(7/44) + 5(2/11) + 6(5/22) \\ = 3.5909$$

- The average roll of the unfair die is 3.5909.
- $E(Y)$ is the mean of the distribution of Y.



Your turn: expected value

z	11	12	13	14	15
$f(z) = P(Z = z)$	0.03	0.03	0.03	0.06	0.26

z	16	17	18	19	20
$f(z) = P(Z = z)$	0.09	0.12	0.20	0.15	0.03

- Calculate $E(Z)$, the expected value of the torque required to loosen the next bolt.

Answer: expected value

$$
\begin{aligned}
E(Z) &= 11(0.03) + 12(0.03) + 13(0.03) + 14(0.06) + 15(0.26) \\
&= 16(0.09) + 17(0.12) + 18(0.20) + 19(0.15) + 20(0.03) \\
&= 16.35
\end{aligned}
$$

- The average torque required to loosen the next bolt is 16.35 units.

Outline

Variance

- **Variance**: the variance $Var(X)$ (also called σ^2) of a random variable X is given by:

$$\mathrm{Var}(X) = \sum_x (x - E(X))^2 f(x)$$

- Shortcut formulas:

$$\begin{aligned}\mathrm{Var}(X) &= \left[\sum_x x^2 f(x)\right] - (E(X))^2 \\ &= E(X^2) - E^2(X)\end{aligned}$$

- The variance is the average squared deviation of random variable from its mean.
- **Standard deviation**: $SD(X) = \sigma = \sqrt{Var(X)}$

Example: calculating the variance

q	1	2	3	7
$f(q)$	0.34	0.1	0.22	0.34

- Long way:

$$\begin{aligned} E(Q) &= 1(0.34) + 2(0.1) + 3(0.22) + 7(0.34) \\ &= 3.58 \\ Var(Q) &= (1-3.58)^2 0.34 + (2-3.58)^2 0.1 \\ &\quad + (3-3.58)^2 0.22 + (7-3.58)^2 0.34 \\ &= 6.56 \end{aligned}$$

- Short way:

$$\begin{aligned} E(Q^2) &= \sum_q q^2 f(q) \\ &= 1(0.34) + 4(0.1) + 9(0.22) + 49(0.34) \\ &= 19.38 \\ Var(Q) &= E(Q^2) - E^2(Q) \\ &= 19.38 - 3.58^2 \\ &= 6.56 \end{aligned}$$

Your turn: calculating the variance

x	1	2	3	4	5	6
$f(x)$	1/6	1/6	1/6	1/6	1/6	1/6

- Calculate $Var(X)$
- Calculate $SD(X)$

Your turn: answers

- $$\begin{aligned} E(X) &= 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) \\ &= 3.5 \\ E(X^2) &= \sum_{x=1}^{6} x^2 f(x) \\ &= 1^2(1/6) + 2^2(1/6) + 3^2(1/6) + 4^2(1/6) + 5^2(1/6) + 6^2(1/6) \\ &= 15.17 \\ Var(X) &= E(X^2) - E^2(X) \\ &= 15.17 - 3.5^2 \\ &= 2.92 \end{aligned}$$
- $SD(X) = \sqrt{2.92} = 1.7088$