Continuous Random Variables: Quantiles, Expected Value, and Variance

Dason Kurkiewicz

Iowa State University

June 11, 2013

Continuous Random Variables: Quantiles, Expected Value, and Variance

Dason Kurkiewicz

Quantiles

Expected Value

Variance

Functions of random variables

Outline

Quantiles

Expected Value

Variance

Functions of random variables

Quantiles of continuous distributions

- The p-quantile of a random variable, X, is the number, $Q(p)$, such that:

$$P(X \leq Q(p)) = p$$

- In terms of the cumulative distribution function (cdf):

$$\begin{aligned} F(Q(p)) &= p \\ Q(p) &= F^{-1}(p) \end{aligned}$$

Example

- Let Y be the time delay (s) between a 60 Hz AC circuit and the movement of a motor on a different circuit.

$$f(y) = \begin{cases} 60 & 0 < y < \frac{1}{60} \\ 0 & \text{otherwise} \end{cases}$$

- $Q(0.95)$:

$$\begin{aligned} 0.95 = P(Y \leq Q(0.95)) &= \int_{-\infty}^{Q(0.95)} f(y)dy \\ &= \int_{-\infty}^{0} 0dx + \int_{0}^{Q(0.95)} 60dy = 0 + (60|_0^{Q(0.95)} \\ &= 60Q(0.95) \\ Q(0.95) &= \frac{0.95}{60} = \frac{19}{1200} \approx 0.0158 \end{aligned}$$

Interpretation: on average, 95% of the time delays will be below 0.0158 seconds.

Example

- You can also calculate quantiles directly from the cdf:

$$F(y) = \begin{cases} 0 & y \leq 0 \\ 60y & 0 < y \leq \frac{1}{60} \\ 1 & y > \frac{1}{60} \end{cases}$$

- $Q(0.25)$:

$$\begin{aligned} 0.25 &= P(Y \leq Q(0.25)) = F(Q(0.25)) \\ &= 60 \cdot Q(0.25) \end{aligned}$$

Hence:

$$Q(0.25) = \frac{0.25}{60} = \frac{1}{240} \approx 0.00417$$

Interpretation: on average, 25% of the time delays will be below 0.00417 seconds.

Your turn: calculating quantiles

- $T \sim \text{Exp}(\alpha = 1/2)$:

$$f(t) = \begin{cases} 0 & t \leq 0 \\ 2e^{-2t} & t \geq 0 \end{cases} \qquad F(t) \begin{cases} 0 & t < 0 \\ 1 - e^{-2t} & t \geq 0 \end{cases}$$

- Find:
  1. $Q(0.05)$
  2. $Q(0.5)$
  3. $Q(p)$ for some p with $0 \leq p \leq 1$

Answers: calculating quantiles

1. $Q(0.05)$:

$$\begin{aligned} 0.05 &= P(T \le Q(0.05)) = F(Q(0.05)) = 1 - e^{-2Q(0.05)} \\ 0.95 &= e^{-2Q(0.05)} \\ \log(0.95) &= -2Q(0.05) \\ Q(0.05) &= \frac{\log(0.95)}{-2} \approx 0.0256 \end{aligned}$$

2. $Q(0.5)$:

$$\begin{aligned} 0.5 &= P(T \le Q(0.5)) = F(Q(0.5)) = 1 - e^{-2Q(0.5)} \\ 0.5 &= e^{-2Q(0.5)} \\ \log(0.5) &= -2Q(0.5) \\ Q(0.5) &= \frac{\log(0.5)}{-2} \approx 0.347 \end{aligned}$$

Answers: calculating quantiles

3. $Q(p)$

$$\begin{aligned} p &= P(T \leq Q(p)) = F(Q(p)) = 1 - e^{-2Q(p)} \\ 1 - p &= e^{-2Q(p)} \\ \log(1-p) &= -2Q(p) \\ Q(p) &= \frac{\log(1-p)}{-2} \end{aligned}$$

Outline

Quantiles

Expected Value

Variance

Functions of random variables

Expected value

- The expected value of a continuous random variable is:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

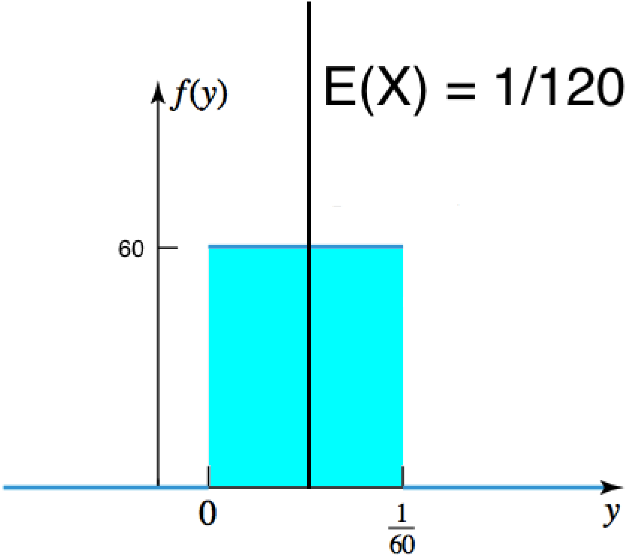
- As with continuous random variables, $E(X)$ (often denoted by μ) is the mean of X, a measure of center.

Example: time delay, Y

$$f(y) = \begin{cases} 60 & 0 \le y \le \frac{1}{60} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y \cdot f(y)dy \\ &= \int_{-\infty}^{0} y \cdot 0dy + \int_{0}^{1/60} y \cdot 60dy + \int_{1/60}^{\infty} y \cdot 0dy \\ &= 0 + \left(\frac{y^2}{2} \cdot 60\right)_0^{1/60} + 0 \\ &= \frac{1}{2}\left(\frac{1}{60}\right)^2 \cdot 60 = \frac{1}{120} \end{aligned}$$

E(X) is the "center of mass" of a distribution



Your turn: calculate E(X)

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\alpha} e^{-x/\alpha} & x \geq 0 \end{cases}$$

1. $X \sim \text{Exp}(3)$
2. $X \sim \text{Exp}(\alpha)$

Answers: Calculate E(X)

1. $X \sim \text{Exp}(3)$:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$
$$= \int_{-\infty}^{0} x \cdot 0dx + \int_{0}^{\infty} x \cdot \frac{1}{3}e^{-x/3}dx$$

integration by parts:

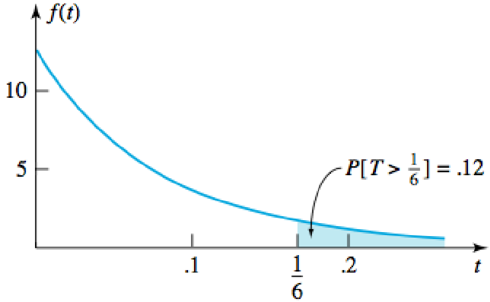
$$= 0 + \left(x(-e^{-x/3})\right)_0^\infty - \int_0^\infty (-e^{-x/3})dx$$
$$= \left(-\infty e^{-\infty/3} + 0e^{-0/3}\right) + \int_0^\infty e^{-x/3}dx$$
$$= 0 + \left(-3e^{-x/3}\right)_0^\infty$$
$$= \left(-3e^{-\infty/3} + 3e^{-0/3}\right)$$
$$= 3$$

2. Similarly, E(X) $= \alpha$ when $X \sim \text{Exp}(\alpha)$.

Example: waiting time for the next student to arrive at the library

- From 12:00 to 12:10 PM, about 12.5 students per minute enter on average.
- Hence, the average waiting time for the next student is $\frac{1}{12.5} = 0.08$ minutes for the next student.
- Let $T \sim \text{Exp}(0.08)$ be the time until the next student arrives.
- P(wait is more than 10 seconds) =

$$P\left(T > 1/6\right) = 1 - F(1/6) = 1 - \left(1 - e^{(-0.08 \cdot 1/6)}\right) = 0.12$$



Outline

Quantiles

Expected Value

Variance

Functions of random variables

Variance

- The variance of a continuous random variable X is:

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)dx$$

Shortcut formulas:

$$\begin{aligned}\text{Var}(X) &= \int_{-\infty}^{\infty} x^2 f(x)dx - E^2(X) \\ &= E(X^2) - E^2(X)\end{aligned}$$

- The standard deviation is $\text{SD}(X) = \sqrt{\text{Var}(X)}$

Your turn: checkout time

Let X denote the amount of time for which a book on 2-hour reserve at a college library is checked out by a randomly selected student and suppose that X has density function

$$f(x) = \begin{cases} .5x & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Calculate:

1. $\mathrm{E}(X)$
2. $\mathrm{Var}(X)$

Answers: checkout time

1.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^2 x \cdot \frac{1}{2}xdx$$
$$= \frac{1}{2}\int_0^2 x^2 dx = \left(\frac{x^3}{6}\right)_0^2 = \frac{8}{6} \approx 1.333$$

2.

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^2 x^2 \frac{1}{2}xdx = \frac{1}{2}\int_0^2 x^3 dx = \left(\frac{x^4}{8}\right)_0^2$$
$$= 2$$

$$Var(X) = E(X^2) - E^2(X) = 2\left(\frac{8}{6}\right)^2$$
$$= \frac{2}{9}$$

Your turn: ecology

- An ecologist wishes to mark off a circular sampling region having radius 10 m. However, the radius of the resulting region is actually a random variable R with pdf:

$$f(r) = \begin{cases} \frac{3}{2}(10 - r)^2 & 9 \le r \le 11 \\ 0 & \text{otherwise} \end{cases}$$

- Calculate:
  1. $E(R)$
  2. $\text{SD}(R)$

Answers: ecology

1.

$$
\begin{aligned}
E(R) &= \int_{-\infty}^{\infty} r \cdot f(r)dr \\
&= \int_{9}^{11} r \cdot \frac{3}{2}(10-r)^2 dr \\
&= \int_{9}^{11} \left(\frac{3}{2}r^3 - 30r^2 + 150r\right) dr \\
&= \left(\frac{3}{8}r^3 - 10r^3 + 75r^2\right)_{9}^{11} \\
&= \left(\frac{3}{8}(11)^3 - 10(11)^3 + 75(11)^2\right) - \left(\frac{3}{8}9^3 - 10(9)^3 + 75(9)^2\right) \\
&= 10
\end{aligned}
$$

Answers: ecology

2.

$$
\begin{aligned}
E(R^2) &= \int_{-\infty}^{\infty} r^2 \cdot f(r)dr \\
&= \int_9^{11} r^2 \cdot \frac{3}{2}(10-r)^2 dr \\
&= \int_9^{11} \left(\frac{3}{2}r^4 - 30r^3 + 150r^2\right) dr \\
&= \left(\frac{3}{10}r^5 - \frac{15}{2}r^4 + 50r^3\right)_9^{11} \\
&= \left(\frac{3}{10}(11)^5 - \frac{15}{2}(11)^4 + 50(11)^3\right) - \left(\frac{3}{10}(9)^5 - \frac{15}{2}(9)^4 + 50(9)^3\right) \\
&= \frac{503}{5} = 100.6
\end{aligned}
$$

$$
\mathrm{Var}(R) = E(R^2) - E^2(R) = \frac{503}{5} - 10^2 = \frac{3}{5} = 0.6
$$

$$
\mathrm{SD}(R) = \sqrt{\mathrm{Var}(R)} = \sqrt{0.6} \approx 2.449
$$

Outline

Quantiles

Expected Value

Variance

Functions of random variables

Expectation of a function of a random variable

- Why does $E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx$?
- It turns out that for any function g of a random variable:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)dx$$

- Hence:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx$$

  if we take $g(X) = X^2$.
- In the ecology example, the expected *area* of the circular sampling region is:

$$E(\pi R^2) = \int_{-\infty}^{\infty} \pi r^2 \cdot f(r)dr$$

  where $\pi R^2 = g(R)$ is the sampling area.

Expectation of a linear function of X

- For constants a and b:

$$
\begin{aligned}
E(aX+b) &= \int_{-\infty}^{\infty} (ax+b)\cdot f(x)dx \\
&= a\underbrace{\int_{-\infty}^{\infty} x\cdot f(x)dx}_{E(X)} + b\underbrace{\int_{-\infty}^{\infty} f(x)dx}_{1} \\
&= \boxed{aE(X)+b}
\end{aligned}
$$

- Example: the expected *diameter* of the ecologist's sampling region is:

$$E(2\cdot R+0) = 2\cdot E(R)+0 = 2\cdot 10 = 20$$

Variance of a linear function of X

- For constants a and b:

$$\begin{aligned}
\text{Var}(aX+b) &= E((aX+b)^2) - E^2(aX+b) \\
&= E(a^2X^2 + abX + b^2) - (aE(X)+b)^2 \\
&= (a^2E(X^2) + abE(X) + b^2) \\
&\quad - (a^2E^2(X) + abE(X) + b^2) \\
&= a^2(E(X^2) - E^2(X)) \\
&= \boxed{a^2 Var(X)}
\end{aligned}$$

- Example: the variance of the *diameter* of the ecologist's sampling region is:

$$\text{Var}(2 \cdot R + 0) = 4\text{Var}(R) = 4 \cdot \frac{503}{5} = \frac{2012}{5}$$

Standardization

- **Standardization**: converting a random variable X into another random variable Z by subtracting the mean and dividing by the standard deviation:

$$Z = \frac{X - E(X)}{SD(X)}$$

- Z has mean 0:

$$\begin{aligned} E(Z) &= E\left(\frac{X - E(X)}{SD(X)}\right) = E\left(\frac{1}{SD(X)} \cdot X - \frac{E(X)}{SD(X)}\right) \\ &= \frac{1}{SD(X)} \cdot E(X) - \frac{E(X)}{SD(X)} = 0 \end{aligned}$$

- Z has variance (and standard deviation) 1:

$$\begin{aligned} \mathrm{Var}(Z) &= \mathrm{Var}\left(\frac{X - E(X)}{SD(X)}\right) = \mathrm{Var}\left(\frac{1}{SD(X)} \cdot X - \frac{E(X)}{SD(X)}\right) \\ &= \frac{1}{SD^2(X)}\mathrm{Var}(X) = \mathrm{Var}(X)\frac{1}{\mathrm{Var}(X)} = 1 \end{aligned}$$