STAT 230
SECTION 2
LECTURE 26

Today's Agenda

Last time:

- Exponential distribution as waiting time in a poisson process
- Mean and variance of the exponential distribution
- Memoryless property

Today (Lec 26, 07/04):

- Brief review exponential distribution
- Percentile
- Generating observations of random variables on a computer

Recall: Exponential distribution

- If $X \sim Exp(\lambda)$, then $f(x) = \lambda \exp(-\lambda x)$ for $x > 0$.

Recall: Exponential distribution

- If $X \sim Exp(\lambda)$, then $f(x) = \lambda \exp(-\lambda x)$ for $x > 0$.
- If $\theta = 1/\lambda$, then we can also write $f(x) = \theta^{-1} \exp(-x/\theta)$ for $x > 0$.

Recall: Exponential distribution

- If $X \sim Exp(\lambda)$, then $f(x) = \lambda \exp(-\lambda x)$ for $x > 0$.
- If $\theta = 1/\lambda$, then we can also write $f(x) = \theta^{-1} \exp(-x/\theta)$ for $x > 0$.
- λ is the rate parameter from Poisson process, and θ is the waiting time parameter derived from it.

Properties of exponential distribution:

- If $X \sim Exp(\theta)$, then $E[X] = \theta$, $Var(X) = \theta^2$.
- If $X \sim Exp(\lambda)$, then $E[X] = 1/\lambda$, $Var(X) = 1/\lambda^2$.

Recall: Exponential distribution

- If $X \sim Exp(\lambda)$, then $f(x) = \lambda \exp(-\lambda x)$ for $x > 0$.
- If $\theta = 1/\lambda$, then we can also write $f(x) = \theta^{-1} \exp(-x/\theta)$ for $x > 0$.
- λ is the rate parameter from Poisson process, and θ is the waiting time parameter derived from it.

Properties of exponential distribution:

- If $X \sim Exp(\theta)$, then $E[X] = \theta,\ Var(X) = \theta^2$.
- If $X \sim Exp(\lambda)$, then $E[X] = 1/\lambda,\ Var(X) = 1/\lambda^2$.
- We can compute them easily with gamma function:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy,\ \alpha > 0.$$

Recall: Exponential distribution

- If $X \sim Exp(\lambda)$, then $f(x) = \lambda \exp(-\lambda x)$ for $x > 0$.
- If $\theta = 1/\lambda$, then we can also write $f(x) = \theta^{-1} \exp(-x/\theta)$ for $x > 0$.
- λ is the rate parameter from Poisson process, and θ is the waiting time parameter derived from it.

Properties of exponential distribution:

- If $X \sim Exp(\theta)$, then $E[X] = \theta,\ Var(X) = \theta^2$.
- If $X \sim Exp(\lambda)$, then $E[X] = 1/\lambda,\ Var(X) = 1/\lambda^2$.
- We can compute them easily with gamma function:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy,\ \alpha > 0.$$

- Exponential distribution is the only memoryless continuous distribution:

$$P(X > s + t | X > s) = P(X > t).$$

Question

Which of the following is true about gamma function?

A $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for $\alpha > 1$

B $\Gamma(\alpha) = (\alpha - 1)!$ for $\alpha \in \mathbb{N}$

C $\Gamma(1/2) = \sqrt{\pi}$

D All of the above

Question

Which of the following is true about gamma function?

A $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for $\alpha > 1$

B $\Gamma(\alpha) = (\alpha - 1)!$ for $\alpha \in \mathbb{N}$

C $\Gamma(1/2) = \sqrt{\pi}$

D All of the above

Question

Suppose the waiting time X until Mukhtar's next bus arrives follows an exponential distribution with parameter $\theta = 1$. What's the waiting time w so that Mukhtar doesn't have to wait longer than w with probability 50%?

Question

Suppose the waiting time X until Mukhtar's next bus arrives follows an exponential distribution with parameter $\theta = 1$. What's the waiting time w so that Mukhtar doesn't have to wait longer than w with probability 50%?

Solution: We know $F(x) = P(X \leq x) = \int_0^x e^{-t}\,\mathrm{d}t = 1 - e^{-x}$ for $x \geq 0$. We want w such that $P(X \leq w) = 0.5$, or

$$F(w) = 0.5 \Leftrightarrow 1 - e^{-w} = 0.5 \Leftrightarrow w = \log(2) \approx 0.693.$$

So with probability 50% Mukhtar won't have to wait longer than $\log(2)$. In other words, $\log(2)$ is the median of the distribution of X.

Definition

The $100 \times q$th percentile (or $100 \times q\%$ quantile) of the distribution of X with cdf F_X is the value c_q, such that

$$F_X(c_q) = q.$$

The quantile is the "inverse" of the cdf.

Definition

The $100 \times q$th percentile (or $100 \times q\%$ quantile) of the distribution of X with cdf F_X is the value c_q, such that

$$F_X(c_q) = q.$$

The quantile is the "inverse" of the cdf.

Remark:

- The probability that X is less than or equal to c_q is (at least) $100 \times q\%$.

Definition

The $100 \times q$th percentile (or $100 \times q\%$ quantile) of the distribution of X with cdf F_X is the value c_q, such that

$$F_X(c_q) = q.$$

The quantile is the "inverse" of the cdf.

Remark:

- The probability that X is less than or equal to c_q is (at least) $100 \times q\%$.
- The **median** of a distribution is its 50% quantile.

Definition

The $100 \times q$th percentile (or $100 \times q\%$ quantile) of the distribution of X with cdf F_X is the value c_q, such that

$$F_X(c_q) = q.$$

The quantile is the "inverse" of the cdf.

Remark:

- The probability that X is less than or equal to c_q is (at least) $100 \times q\%$.
- The **median** of a distribution is its 50% quantile.
- This is different from the **mode**, which is the most likely value, and different from the **mean** which is the expectation (long term "average").

Example

Suppose $X \sim Exp(\theta = 5)$, calculate the 25th percentile and the median of the distribution of X.

Example

Suppose $X \sim Exp(\theta = 5)$, calculate the 25th percentile and the median of the distribution of X. Solution

- $X \sim Exp(\theta = 5)$ has pdf $f(x) = (1/5)\exp(-x/5)$ for $x > 0$.

Example

Suppose $X \sim Exp(\theta = 5)$, calculate the 25th percentile and the median of the distribution of X. Solution

- $X \sim Exp(\theta = 5)$ has pdf $f(x) = (1/5)\exp(-x/5)$ for $x > 0$.
- The cdf of X is then $F(x) = \int_{-\infty}^{x} f(t)\, \mathrm{d}t = \int_0^x (1/5)\exp(-t/5)\, \mathrm{d}t = 1 - \exp(-x/5)$.

Example

Suppose $X \sim Exp(\theta = 5)$, calculate the 25th percentile and the median of the distribution of X. Solution

- $X \sim Exp(\theta = 5)$ has pdf $f(x) = (1/5)\exp(-x/5)$ for $x > 0$.
- The cdf of X is then $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_0^x (1/5)\exp(-t/5)\,\mathrm{d}t = 1 - \exp(-x/5)$.
- Solving $F(c_q) = q$ for x where $q \in (0,1)$ gives

$$F(c_q) = q \Leftrightarrow 1 - q = \exp(-c_q/5) \Leftrightarrow c_q = -5\log(1-q)$$

Example

Suppose $X \sim Exp(\theta = 5)$, calculate the 25th percentile and the median of the distribution of X. Solution

- $X \sim Exp(\theta = 5)$ has pdf $f(x) = (1/5)\exp(-x/5)$ for $x > 0$.
- The cdf of X is then
  $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{0}^{x} (1/5)\exp(-t/5)\,\mathrm{d}t = 1 - \exp(-x/5)$.
- Solving $F(c_q) = q$ for x where $q \in (0, 1)$ gives

$$F(c_q) = q \Leftrightarrow 1 - q = \exp(-c_q/5) \Leftrightarrow c_q = -5\log(1-q)$$

- For $q = 0.25$ we get $c_{0.25} = -5\log(0.75) \approx 1.438$ as the 25th percentile.

Example

Suppose $X \sim Exp(\theta = 5)$, calculate the 25th percentile and the median of the distribution of X. Solution

- $X \sim Exp(\theta = 5)$ has pdf $f(x) = (1/5)\exp(-x/5)$ for $x > 0$.
- The cdf of X is then $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{0}^{x}(1/5)\exp(-t/5)\,\mathrm{d}t = 1 - \exp(-x/5)$.
- Solving $F(c_q) = q$ for x where $q \in (0, 1)$ gives

$$F(c_q) = q \Leftrightarrow 1 - q = \exp(-c_q/5) \Leftrightarrow c_q = -5\log(1 - q)$$

- For $q = 0.25$ we get $c_{0.25} = -5\log(0.75) \approx 1.438$ as the 25th percentile.
- For $q = 0.5$ we get $c_{0.5} = -5\log(0.5) \approx 3.466$ as the median.

Question

Suppose X and Y are continuous random variables satisfying $P(X \leq x) < P(Y \leq x)$. Let s_x and s_y denote the 50th percentiles of the distributions of X and Y, respectively. Then

A $s_x < s_y$

B $s_y < s_x$

C It is possible that $s_x = s_y$

D Unable to determine a relationship between s_x and s_y.

Suppose X and Y are continuous random variables satisfying $P(X \leq x) < P(Y \leq x)$. Let s_x and s_y denote the 50th percentiles of the distributions of X and Y, respectively. Then

A $s_x < s_y$

B $s_y < s_x$

C It is possible that $s_x = s_y$

D Unable to determine a relationship between s_x and s_y.

Suppose X and Y are continuous random variables satisfying $P(X \leq x) < P(Y \leq x)$. Let s_x and s_y denote the 50th percentiles of the distributions of X and Y, respectively. Then

A $s_x < s_y$

B $s_y < s_x$

C It is possible that $s_x = s_y$

D Unable to determine a relationship between s_x and s_y.

Solution: Note that $F_X(s_x) = 0.5 = F_Y(s_y)$. By assumption,

$$0.5 = F_X(s_x) < F_Y(s_x).$$

Since F_Y is increasing, we must have $s_y < s_x$.

Sampling realizations of random variables

- It is "easy" to sample from the continuous uniform distribution $\mathit{Unif}(0,1)$ on a computer, that is, we can easily generate numbers that look like they come from $\mathit{Unif}(0,1)$.

Sampling realizations of random variables

- It is "easy" to sample from the continuous uniform distribution $\mathit{Unif}(0,1)$ on a computer, that is, we can easily generate numbers that look like they come from $\mathit{Unif}(0,1)$.
  - In R: `uniforms <- runif(n)` creates a vector of length `n` with realizations of $\mathit{Unif}(0,1)$.

Sampling realizations of random variables

- It is “easy” to sample from the continuous uniform distribution $\mathit{Unif}\,(0,1)$ on a computer, that is, we can easily generate numbers that look like they come from $\mathit{Unif}\,(0,1)$.
  - In R: `uniforms <- runif(n)` creates a vector of length `n` with realizations of $\mathit{Unif}\,(0,1)$.
- But how can we sample from any distribution?

Sampling realizations of random variables

- It is "easy" to sample from the continuous uniform distribution $\mathit{Unif}(0,1)$ on a computer, that is, we can easily generate numbers that look like they come from $\mathit{Unif}(0,1)$.
  - In R: `uniforms <- runif(n)` creates a vector of length `n` with realizations of $\mathit{Unif}(0,1)$.
- But how can we sample from any distribution?
- If F is a continuous and strictly increasing cdf of some random variable X and if $U \sim \mathit{Unif}(0,1)$, then the random variable $Y = F^{-1}(U)$ has cdf F:

Sampling realizations of random variables

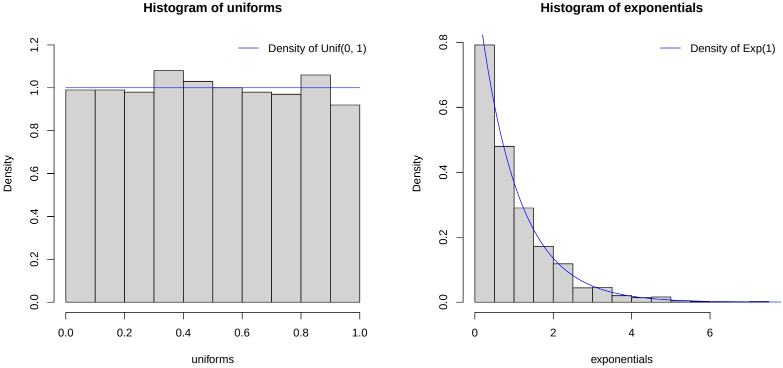
- It is "easy" to sample from the continuous uniform distribution $Unif(0, 1)$ on a computer, that is, we can easily generate numbers that look like they come from $Unif(0, 1)$.
  - In R: `uniforms <- runif(n)` creates a vector of length `n` with realizations of $Unif(0, 1)$.
- But how can we sample from any distribution?
- If F is a continuous and strictly increasing cdf of some random variable X and if $U \sim Unif(0, 1)$, then the random variable $Y = F^{-1}(U)$ has cdf F:

$$F_Y(y) = P(F^{-1}(U) \leq y) = P(U \leq F(y)) = F(y), \quad y \in \mathbb{R}.$$

- So for continuous random variables with strictly increasing cdf, we have $Y = F^{-1} \sim F$. Let look at an example.

Example

We sample `uniforms <- runif(1000)` and then `exponentials <- -log(1-uniforms)`.



Generalization

More generally, one can show the following theorem (see p172).

Theorem

Let F be an arbitrary cdf and $U \sim Unif(0,1)$. The random variable $X = F^{-1}(U)$ has cdf F, where

$$F^{-1}(y) = \inf_{x \in \mathbb{R}} \{F(x) \geq y\}.$$

Example

A random variable X has a $\text{Burr}(c, k)$ distribution, if X has pdf

$$f(x) = ckx^{c-1}(1 + x^c)^{-k-1}, \quad x > 0,$$

where c, k are parameters.

a) Show that f is a valid pdf.
b) Derive a function g such that $g(U)$ has pdf f if $U \sim Unif(0, 1)$.

a) Note that $f(x) \geq 0$ for all x and

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_0^{\infty} ckx^{c-1}(1+x^c)^{-k-1}\,\mathrm{d}x = \left[-(1+x^c)^{-k}\right]_0^{\infty} = 1.$$

b) We first compute the cdf as

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_0^{x} ckt^{c-1}(1+t^c)^{-k-1}\,\mathrm{d}t = 1-(1+x^c)^{-k}$$

for $x > 0$ and 0 otherwise. From the theorem, we know we have to choose $g(u) = F^{-1}(u)$. Solving $F(x) = u$ for x gives

$$g(u) = F^{-1}(u) = \left((1-u)^{-1/k} - 1\right)^{1/c}.$$