Chapter 5 part 2

Continuous Random Variables

Jingchen (Monika) Hu

Vassar College

MATH 241

Outline

Outline

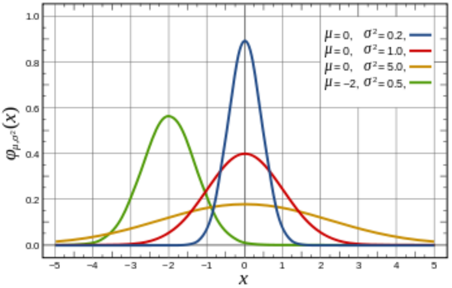
Normal Distribution

Definition

A continuous random variable X has a Normal distribution distribution with mean μ and variance σ^2 if its pdf is

$$X \sim \mathsf{N}(\mu, \sigma^2) \iff f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ \textit{where } x \in \mathbb{R}$$

Pdf: unimodal and symmetric, bell shaped random variable



The normal pdf is well-defined

(not required)

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

Question

Use the fact that normal pdf is well define, calculate the integral

$$\int_{-\infty}^{\infty} e^{-\frac{(x+2)^2}{6}} dx = ?$$

The normal pdf is well-defined

(not required)

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

Question

Use the fact that normal pdf is well define, calculate the integral

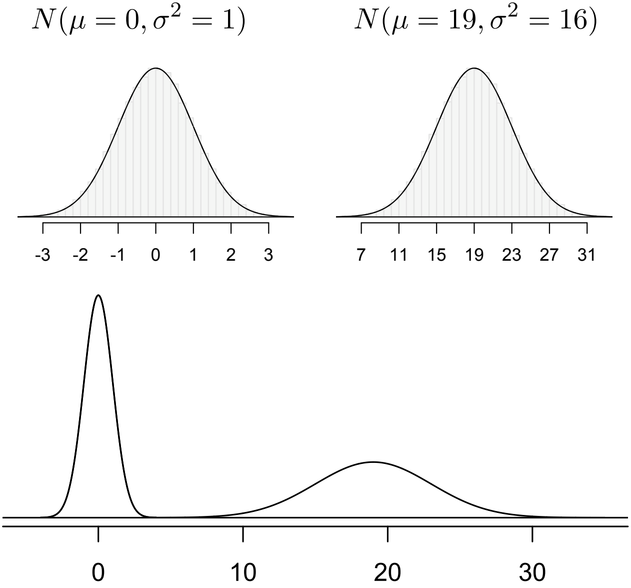
$$\int_{-\infty}^{\infty} e^{-\frac{(x+2)^2}{6}} dx = ?$$

Suppose we have a random variable $X \sim \mathsf{N}(\mu = -2, \sigma^2 = 3)$, then

$$1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot 3}}\, e^{-\frac{(x+2)^2}{6}} dx$$

$$\Longrightarrow \int_{-\infty}^{\infty} e^{-\frac{(x+2)^2}{6}} dx = \sqrt{6\pi}$$

Normal distributions with different parameters



Mean and variance of Normal random variable

If $X \sim \mathsf{N}(\mu, \sigma^2)$, and $Z = \frac{X-\mu}{\sigma}$, then Z has a standard normal distribution (more later in Chapter 5.7).

$$Z \sim \mathsf{N}(0, 1)$$

- $E[X] = \mu$ (required)

Since $X \sim \mathsf{N}(\mu, \sigma^2) \iff X = \sigma Z + \mu$ and $Z \sim \mathsf{N}(0, 1)$, it's suffice to show the mean and variance of standard normal are 0 and 1.

$$\begin{aligned} E[Z] &= \int_{-\infty}^{\infty} x f_Z(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \ e^{-\frac{x^2}{2}} dx \\ &= -\frac{1}{\sqrt{2\pi}} \ e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} = 0 \end{aligned}$$

Mean and variance of Normal random variable

If $X \sim \mathsf{N}(\mu, \sigma^2)$, and $Z = \frac{X-\mu}{\sigma}$, then Z has a standard normal distribution (more later in Chapter 5.7).

$$Z \sim \mathsf{N}(0, 1)$$

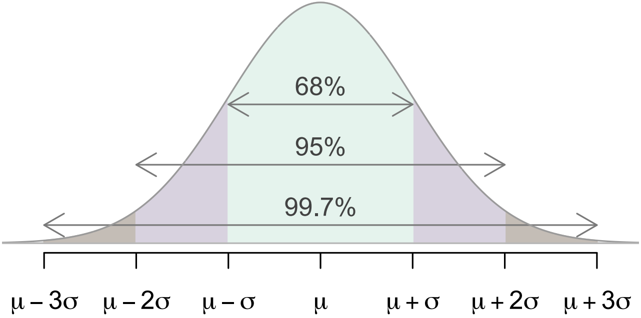
- $Var(X) = \sigma^2$ (required)

  Since $X \sim \mathsf{N}(\mu, \sigma^2) \Longleftrightarrow X = \sigma Z + \mu$ and $Z \sim \mathsf{N}(0, 1)$, it's suffice to show the mean and variance of standard normal are 0 and 1.

$$\begin{aligned}
Var[Z] &= E[Z^2] - (E[Z])^2 = E[Z^2] = \int_{-\infty}^{\infty} x^2 f_Z(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \, e^{-\frac{x^2}{2}} dx \ \ (\text{integration by parts} : u = x, dv = xe^{-x^2/2} dx) \\
&= \frac{1}{\sqrt{2\pi}} \left\{ - xe^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right\} \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1
\end{aligned}$$

68-95-99.7 Rule

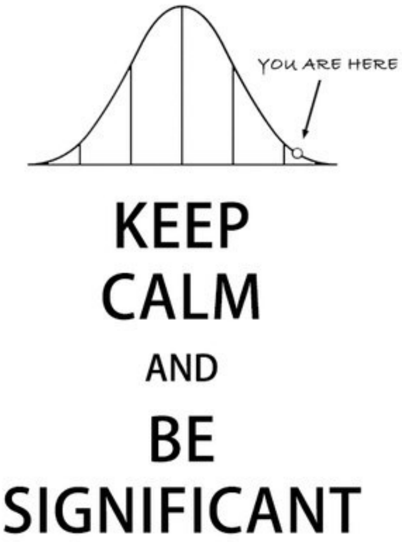
- A random variable X has a normal distribution,
  - about 68% probability X falls within 1 SD of the mean,
  - about 95% probability X falls within 2 SD of the mean,
  - about 99.7% probability X falls within 3 SD of the mean.
- The probability of X falls 4, 5, or more standard deviations away from the mean is very low.



Normal probability calculation

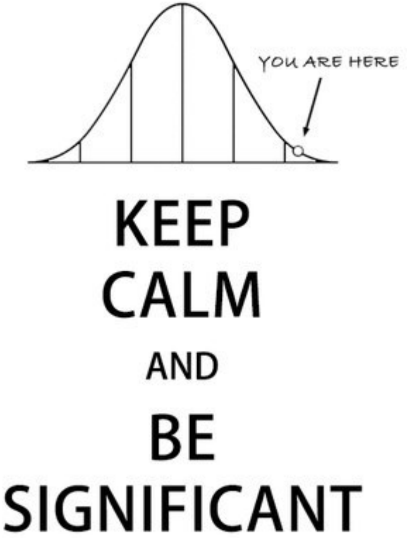
- We denote $\phi(x)$ and $\Phi(x)$ as pdf and cdf of the standard normal distribution respectively.
- Probability calculations for X in terms of Z:

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$



Question

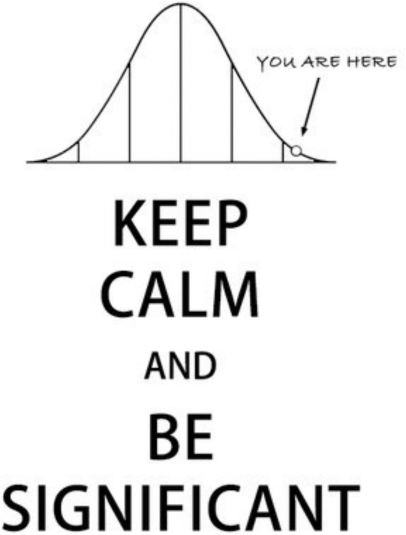
Using the 68-95-99.7 rule, find the standard normal cdf $\Phi(-2)$ and $\Phi(2)$



Question

Using the 68-95-99.7 rule, find the standard normal cdf $\Phi(-2)$ and $\Phi(2)$

$$\begin{aligned}\Phi(-2) =& P(Z < -2) \\ &= \frac{1 - P(-2 < Z < 2)}{2} \\ &= \frac{1 - 0.95}{2} = 0.025\end{aligned}$$



Question

Using the 68-95-99.7 rule, find the standard normal cdf $\Phi(-2)$ and $\Phi(2)$

$$\begin{aligned}\Phi(-2) =& P(Z < -2) \\ =& \frac{1 - P(-2 < Z < 2)}{2} \\ =& \frac{1 - 0.95}{2} = 0.025\end{aligned}$$

$$\Phi(2) = 1 - \Phi(-2) = 0.975$$

Question

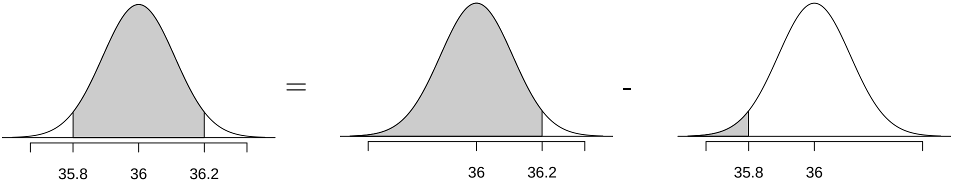
At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle will fails the quality control inspection. What percent of bottles pass the quality control inspection?

Question

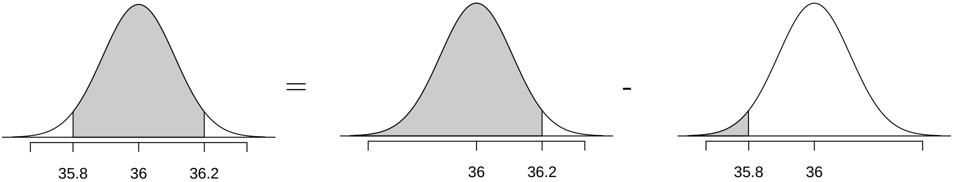
At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle will fails the quality control inspection. What percent of bottles pass the quality control inspection?

$$X \sim \mathsf{N}(36, 0.11^2), \quad P(35.8 < X < 36.2) =?$$

$$X \sim \mathsf{N}(36, 0.11^2), \quad P(35.8 < X < 36.2) = ?$$



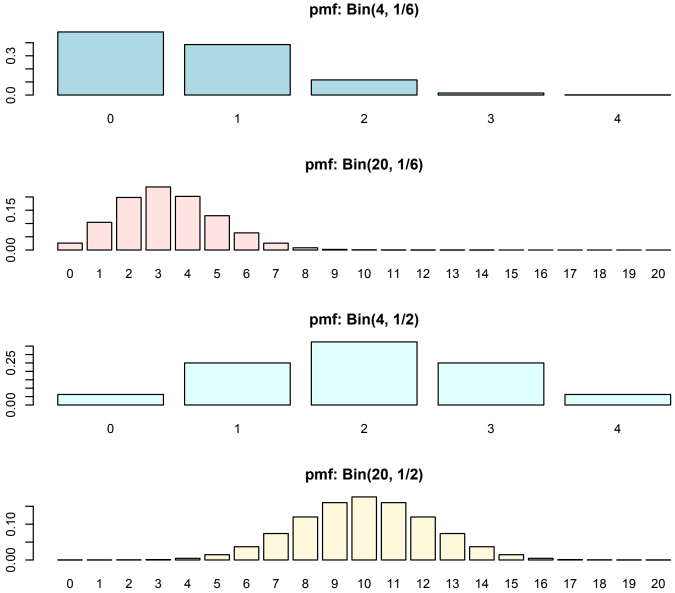
$$X \sim \mathsf{N}(36, 0.11^2), \quad P(35.8 < X < 36.2) = ?$$



$$z_1 = \frac{35.8 - 36}{0.11} = -1.82, \quad z_2 = \frac{36.2 - 36}{0.11} = 1.82$$

$$\begin{aligned} P(35.8 < X < 36.2) &= P(-1.82 < Z < 1.82) \\ &= P(Z < 1.82) - P(Z < -1.82) \\ &= 0.9656 - 0.0344 = 0.9312 \end{aligned}$$

Normal approximation to the Binomial distribution



Normal approximation to the Binomial distribution

Let $X \sim \text{Bin}(n, p)$. When n is large enough, or more specifically, $np(1-p) \geq 10$, the binomial distribution can be approximated by the normal distribution

$$P(X = i) \approx P(i - 0.5 < Y < i + 0.5), \quad Y \sim \text{N}(\mu, \sigma^2)$$

with parameters $\mu = np$ and $\sigma^2 = np(1-p)$.

- Probability calculation (actually, approximation)

$$\begin{aligned}
P(a \leq X \leq b) &\approx P(a - 0.5 < Y < b + 0.5) \\
&= P\left(\frac{a - 0.5 - np}{\sqrt{np(1-p)}} < \frac{Y - \mu}{\sigma} < \frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right) \\
&= \Phi\left(\frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 0.5 - np}{\sqrt{np(1-p)}}\right)
\end{aligned}$$

Question

A recent study found that "Facebook users get more than they give". For example:

- Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content "liked" an average of 20 times
- 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

This is because there are "power users" who contribute much more content than the typical user. The same study found that approximately 25% of Facebook users are considered power users. It also found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

Question

A recent study found that "Facebook users get more than they give". For example:

- Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content "liked" an average of 20 times
- 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

This is because there are "power users" who contribute much more content than the typical user. The same study found that approximately 25% of Facebook users are considered power users. It also found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

- We are given that $X \sim \text{Bin}(n = 245, p = 0.25)$, and we are asked for the probability

$$P(X \geq 70) = p(70) + p(71) + \cdots + p(245) = 1 - p(0) - p(1) - \cdots - p(69)$$

Question

A recent study found that "Facebook users get more than they give". For example:

- Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content "liked" an average of 20 times
- 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

This is because there are "power users" who contribute much more content than the typical user. The same study found that approximately 25% of Facebook users are considered power users. It also found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

- We are given that $X \sim \text{Bin}(n = 245, p = 0.25)$, and we are asked for the probability

$$P(X \geq 70) = p(70)+p(71)+\cdots+p(245) = 1-p(0)-p(1)-\cdots-p(69)$$

- To use normal approximation, first check conditions

$$np(1-p) = 245 \times 0.25 \times 0.75 = 45.94 \geq 10$$

Question

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

Use Normal approximation.

$$\mu = 245 \times 0.25 = 61.25, \quad \sigma = \sqrt{245 \times 0.25 \times 0.75} = 6.78$$

$$Y \sim \mathsf{N}(\mu, \sigma^2), \quad P(X \geq 70) \approx P(Y > 70 - 0.5)$$

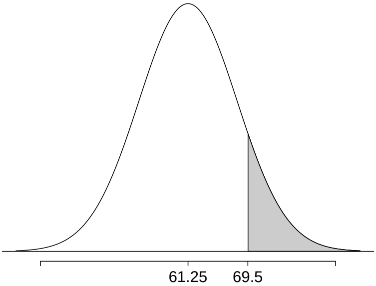
Question

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

Use Normal approximation.

$$\mu = 245 \times 0.25 = 61.25, \quad \sigma = \sqrt{245 \times 0.25 \times 0.75} = 6.78$$

$$Y \sim \mathsf{N}(\mu, \sigma^2), \quad P(X \geq 70) \approx P(Y > 70 - 0.5)$$



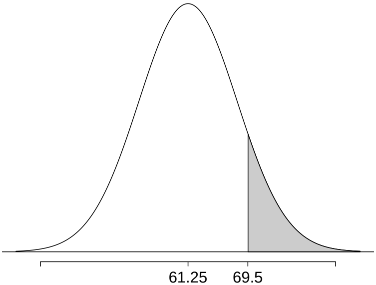
Question

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

Use Normal approximation.

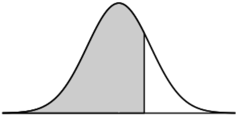
$$\mu = 245 \times 0.25 = 61.25, \quad \sigma = \sqrt{245 \times 0.25 \times 0.75} = 6.78$$

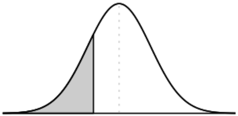
$$Y \sim \mathsf{N}(\mu, \sigma^2), \quad P(X \geq 70) \approx P(Y > 70 - 0.5)$$



$$\begin{aligned} P(X \geq 70) &\approx P(Y > 70 - 0.5) \\ &= P\left(Z > \frac{69.5 - 61.25}{6.78}\right) \\ &= P(Z > 1.22) \\ &= 1 - 0.8888 = 0.1112 \end{aligned}$$

Recap: Normal distribution $X \sim \mathsf{N}(\mu, \sigma^2)$





- Mean μ, variance σ^2
- Symmetry

$$f(\mu - x) = f(\mu + x), F(\mu - x) = 1 - F(\mu + x)$$

- Standard normal distribution $\mu = 0, \sigma^2 = 1$.

$$X \sim \mathsf{N}(\mu, \sigma^2) \Longleftrightarrow Z = \frac{X - \mu}{\sigma} \sim \mathsf{N}(0, 1)$$

- Find probability using $\Phi(\cdot)$ table

$$P(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

- Normal approximation to Binomial $X \sim \mathsf{Bin}(n, p)$

$$Y \sim \mathsf{N}(\mu = np, \sigma^2 = np(1 - p))$$

$$P(X = i) \approx P(i - 0.5 < Y < i + 0.5)$$

Review: continuous distributions

Name	Range	pdf $f(x)$	mean	variance
Unif(α, β)	$[\alpha, \beta]$	$\frac{1}{\beta-\alpha}$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$
N(μ, σ^2)	$(-\infty, \infty)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$	μ	σ^2

Outline

A (important!) theorem on finding pdf of $g(X)$

Suppose X is a continuous random variable with pdf $f_X(x)$. If a function $g(x)$ is

1. monotonic (increasing or decreasing), and
2. differentiable (and thus continuous),

then the random variable defined by $Y = g(X)$ has pdf

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

Or more rigorously,

$$f_Y(y) = \begin{cases} f_X\left[g^{-1}(y)\right] \cdot \left| \frac{d}{dy} g^{-1}(y) \right| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

Question

Let $X \sim \mathsf{N}(\mu, \sigma^2)$, what distribution does $Y = (X - \mu)/\sigma$ have?

Question

Let $X \sim \mathsf{N}(\mu, \sigma^2)$, what distribution does $Y = (X - \mu)/\sigma$ have?

In order to use the previous theorem, follow these steps:

1. Check if $g(x) = (x - \mu)/\sigma$ is monotonic and differentiable
2. Compute the inverse function $y = g(x) \Longleftrightarrow x = g^{-1}(y)$
3. Compute the derivative $\frac{dx}{dy}$
4. Identify the range of the new random variable $Y = g(X)$.
5. Apply the formula

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

Question

Let $X \sim \mathrm{N}(\mu, \sigma^2)$, what distribution does $Y = (X - \mu)/\sigma$ have?

In order to use the previous theorem, follow these steps:

1. Check if $g(x) = (x - \mu)/\sigma$ is monotonic and differentiable ✓
2. Compute the inverse function $y = g(x) \iff x = g^{-1}(y)$

$$x = \sigma y + \mu$$

3. Compute the derivative $\frac{dx}{dy} = \sigma$
4. Identify the range of the new random variable $Y = g(X)$. $y \in \mathbb{R}$
5. Apply the formula

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot |\sigma|$$
$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\sigma y+\mu-\mu)^2}{2\sigma^2}} \cdot \sigma = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Formula cannot be used. So two steps: (1) write the cdf $F_Y(y)$ in the form of $F_X(\cdot)$.

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Formula cannot be used. So two steps: (1) write the cdf $F_Y(y)$ in the form of $F_X(\cdot)$.

2) Take derivative to get pdf $f_Y(y) = \frac{d}{dy} F_Y(y)$.

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Formula cannot be used. So two steps: (1) write the cdf $F_Y(y)$ in the form of $F_X(\cdot)$. For any $y \geq 0$,

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y)$$

2) Take derivative to get pdf $f_Y(y) = \frac{d}{dy} F_Y(y)$.

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Formula cannot be used. So two steps: (1) write the cdf $F_Y(y)$ in the form of $F_X(\cdot)$. For any $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \end{aligned}$$

2) Take derivative to get pdf $f_Y(y) = \frac{d}{dy} F_Y(y)$.

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Formula cannot be used. So two steps: (1) write the cdf $F_Y(y)$ in the form of $F_X(\cdot)$. For any $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

2) Take derivative to get pdf $f_Y(y) = \frac{d}{dy} F_Y(y)$.

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Formula cannot be used. So two steps: (1) write the cdf $F_Y(y)$ in the form of $F_X(\cdot)$. For any $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

2) Take derivative to get pdf $f_Y(y) = \frac{d}{dy} F_Y(y)$.

$$f_Y(y) = \frac{d}{dy} \left[F_X(\sqrt{y}) - F_X(-\sqrt{y})\right]$$

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Formula cannot be used. So two steps: (1) write the cdf $F_Y(y)$ in the form of $F_X(\cdot)$. For any $y \geq 0$,

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\
&= P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})
\end{aligned}
$$

2) Take derivative to get pdf $f_Y(y) = \frac{d}{dy} F_Y(y)$.

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(-\sqrt{y})\right] \\
&= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} - f_X(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right)
\end{aligned}
$$

Question

Let X have a standard normal distribution. What is the pdf of $Y = X^2$?

Formula cannot be used. So two steps: (1) write the cdf $F_Y(y)$ in the form of $F_X(\cdot)$. For any $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

2) Take derivative to get pdf $f_Y(y) = \frac{d}{dy} F_Y(y)$.

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(-\sqrt{y})\right] \\ &= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} - f_X(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right) \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{(\sqrt{y})^2}{2}}\frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}}e^{-\frac{(-\sqrt{y})^2}{2}}\frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{y}{2}} \end{aligned}$$