Special Continuous random variables

March 27, 2019

Random Variables (RV)

- Given a probability space (Ω, P)
- A Random variable is a function

$$X : \Omega \to \mathbb{R}$$

- If Range(X) is finite or countable: X is discrete RV
- If Range(X) is uncountable (contains interval): X is continuous RV
- Use RV to model real random quantity

Probability of RV

- Suppose X is a RV and B is a subset in Range(X)
- $X^{-1}(B)$= set of all points ω in Ω so that $X(\omega)$ in B
- $X^{-1}(B)$ is an event, can find $P(X^{-1}(B))$
- Write: $P(X \in B) = P(X^{-1}(B))$
- $P(X \in B)$: Probability that the values of X are in B

Probability Density Function

- X : continuous Random Variable (RV)
- There exists a continuous positive function $f_X(x)$ such that:
    - Domain of f_X = Range of X
    - For all subsets B of Range(X)

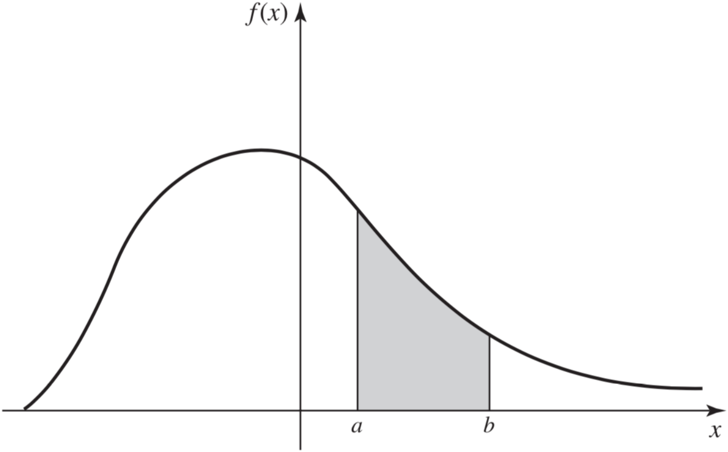
$$P(X \in B) = \int_B f_X(x)dx$$

- $f_X(x)$ is called the Probability density function (pdf) of X

Properties

- X is determined by its pdf $f_X(x)$
- Know $f_X \to$ know everything about X
- $$P(X \leq a) = P(X \in (-\infty, a)) = \int_{-\infty}^{a} f_X(x)dx$$
- $$\int_{-\infty}^{\infty} f_X(x)dx = P(X \in (-\infty, \infty)) = 1$$

Graph of pdf



Expectation

- Expectation of continuous RV

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

- Expectation of a function of RV

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

- Variance of continuous RV

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

Uniform Random variables

Definition

X is uniform RV on $[a, b]$ if its pdf is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

Any value in $[a, b]$ is equally likely to be value of X.
Denote $X \sim \text{Uni}[a, b]$.
If a quantity modeled by Uniform RV, then say that is is *uniformly distributed*.

Definition

X is uniform RV on $[a, b]$ if its pdf is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

Any value in $[a, b]$ is equally likely to be value of X.
Denote $X \sim \text{ Uni}[a, b]$.
If a quantity modeled by Uniform RV, then say that is is *uniformly distributed*.

Uniform pdf

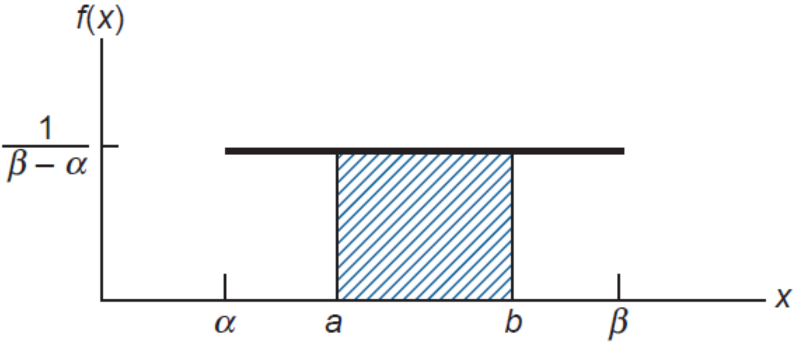


Figure: Pdf of Uni$[a, b]$

Calculating Probability

If $a < c < d < b$ then

$$P(c < X < d) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$$

Example

Buses arrive at a specified stop at 15-minute intervals starting at 7 A.M. That is, they arrive at 7, 7:15, 7:30, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits

(a) less than 5 minutes for a bus

(b) at least 12 minutes for a bus.

Solution

- X = number of minutes after 7 of arrival time $\sim$ Uni$[0, 30]$
- P(wait less than 5 minutes)
  $= P(10 < X < 15) + P(25 < X < 30)$
  $= 5/30 + 5/30 = 1/3$
- P(wait more than 12 minutes)
  $= P(0 < X < 3) + P(25 < X < 28)$
  $= 3/30 + 3/30 = 1/5$

Solution

- X = number of minutes after 7 of arrival time $\sim$ Uni$[0, 30]$
- P(wait less than 5 minutes)
  $= P(10 < X < 15) + P(25 < X < 30)$
  $= 5/30 + 5/30 = 1/3$
- P(wait more than 12 minutes)
  $= P(0 < X < 3) + P(25 < X < 28)$
  $= 3/30 + 3/30 = 1/5$

Expectation

- Expectation of Uniform RV:
  $X \sim \text{Uni}[a, b]$

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

Variance

- $$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx$$
  $$= \frac{a^2 + ab + b^2}{3}$$
- Variance:
  $$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{(b-a)^2}{12}$$

Normal random variables

Definition

Continuous RV X is said to be normally distributed with parameter μ and σ^2 if its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

Denote $X \sim \mathcal{N}(\mu, \sigma^2)$.

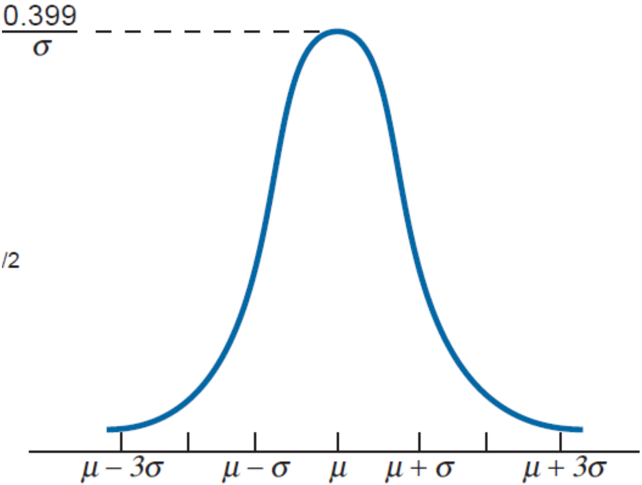


Figure: PDF of $\mathcal{N}(\mu, \sigma^2)$

Moments

- $E(X) = \mu$
- $\mathrm{Var}(X) = \sigma^2$

Application

- Normal distibution is the most widely used distribution
- Many random phenomena obey a normal distribution
- Ex: the height and weight of a person, accuracy of shots from a gun...

Meaning

- Normal RV is used to approximate Binomial (n, p) when n is large
- can use $\mathcal{N}(np, np(1-p))$ instead
- Usually if np and $np(1-p)$ both larger than 5 then approximation is good.

Meaning

- Normal RV is used to approximate Binomial (n, p) when n is large
- can use $\mathcal{N}(np, np(1-p))$ instead
- Usually if np and $np(1-p)$ both larger than 5 then approximation is good.

Central limit theorems

- If $X_1, \ldots, X_n$ are independently identical distributed RVs with mean μ and variance σ^2, and

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

- then $\sqrt{n}(\bar{X} - \mu) \sim \mathcal{N}(0, \sigma^2)$.

Standard normal

- To calculate probability of normal RV we need a special RV:
- Let $Z \sim \mathcal{N}(0, 1)$
- Put $\Phi(x) = P(Z < x)$ for any x
- If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

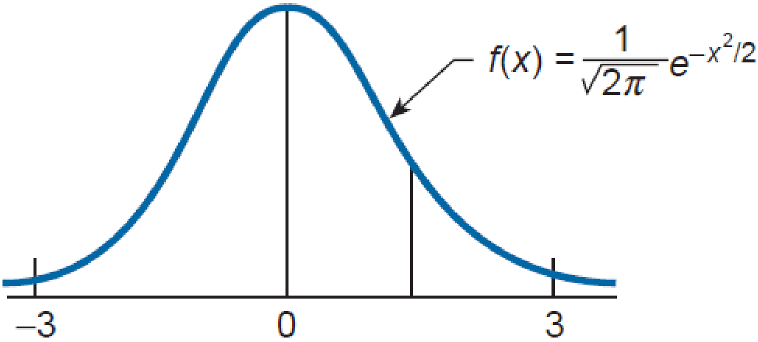


Figure: Pdf of Z

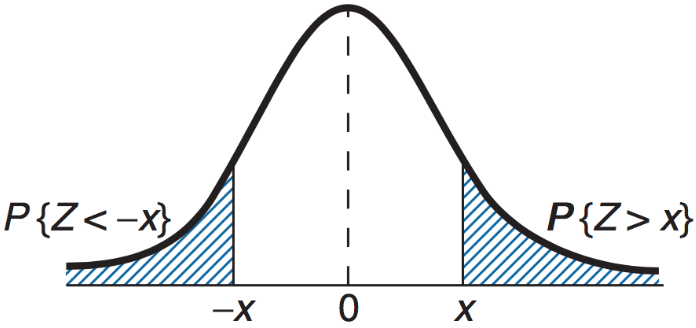
Calculate $N(\mu, \sigma^2)$

Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$. Then

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

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

Symmetry



Negative x

If $x > 0$ then

$$\begin{aligned}\Phi(-x) &= P(Z < -x) \\ &= P(Z > x) \\ &= 1 - P(Z < x) \\ &= 1 - \Phi(x)\end{aligned}$$

We only need to know value of $\Phi(x)$ for $x > 0$

Table of $\Phi(x)$

TABLE AI *Standard Normal Distribution Function:* $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Example

If X is a normal random variable with mean $\mu = 3$ and variance $\sigma^2 = 16$, find

(a) $P(X < 11)$

(b) $P(X > -1)$

(c) $P(2 < X < 7)$.

Solution

$$P\{X < 11\} = P\left\{\frac{X-3}{4} < \frac{11-3}{4}\right\}$$

$$= \Phi(2)$$

$$= .9772$$

$$P\{X > -1\} = P\left\{\frac{X-3}{4} > \frac{-1-3}{4}\right\}$$

$$= P\{Z > -1\}$$

$$= P\{Z < 1\}$$

$$= .8413$$

$$
\begin{aligned}
P\{2 < X < 7\} &= P\left\{\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right\} \\
&= \Phi(1) - \Phi(-1/4) \\
&= \Phi(1) - (1 - \Phi(1/4)) \\
&= .8413 + .5987 - 1 = .4400 \quad \blacksquare
\end{aligned}
$$

Binomial approximation

- Suppose $Y \sim$ Bino(n, p) where n is large and np is not too small
- Y can be approximated by $X \sim \mathcal{N}(np, np(1-p))$
- Y is discrete, X is continuous
- so we have to "fill the gap"

$$P(Y = i) \approx P(i - \frac{1}{2} < X < i + \frac{1}{2})$$

Example

Toss a fair coin 40 times. Y is the number of heads. Calculate $P(Y = 20)$ using normal approximation and direct computation.

Solution

Approximate Y by $X \sim \mathcal{N}(20, 10)$

$$
\begin{aligned}
P(Y = 20) &\approx P(19.5 < X < 20.5) \\
&= P\left(\frac{19.5 - 20}{\sqrt{10}} < Z < \frac{20.5 - 20}{\sqrt{10}}\right) \\
&= \Phi(.16) - \Phi(-.16) \\
&= 2\Phi(.16) - 1 \\
&= .1272
\end{aligned}
$$

Exact value

$$P(Y = 20) = \binom{40}{20}(.5)^{40} = .1254$$

Example

The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that, on the average, only 30 percent of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that more than 150 first-year students attend this college.

Solution

X: number of attending students

$X \sim \text{Bino}(450, .3) \approx \mathcal{N}(135, 94.5)$

$$
\begin{aligned}
P(X > 150) &= P(X \geq 150.5) \\
&\approx P(Z \geq \frac{150.5 - 135}{\sqrt{94.5}}) \\
&= 1 - \Phi(1.59) \\
&= .0559 \approx 5.6\%
\end{aligned}
$$

Example

Let $Y = 3X^2$ and $X \sim \mathcal{N}(6, 1)$, find

(a) $E(Y)$

(b) $P(Y > 120)$

Solution

(a)

$$\begin{aligned} E(3X^2) &= 3E(X^2) \\ &= 3(\operatorname{Var}(X) + E(X)^2) \\ &= 3(1+36) = 111 \end{aligned}$$

(b)

$$P(3X^2 > 120) = P(X^2 > 40)$$
$$= P(X > \sqrt{40}) + P(X < -\sqrt{40})$$
$$= P(Z > \frac{\sqrt{40} - 6}{1}) + P(Z < \frac{-\sqrt{40} - 6}{1})$$
$$= (1 - \Phi(.3246)) + \Phi(-12.32) = .3727 + 0$$

Practice

Suppose that the height, in cm, of a 25-year-old man is a normal random variable with parameters $\mu = 170$ and $\sigma^2 = 6.25$. What percentage of 25- year-old men are over 167 cm tall? What percentage of men in the 175cm club are over 180cm?

Sum of normal RVs

- $X_1, \ldots, X_n$ independent
- $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$
- $X = \sum_{i=1}^{n} X_i$
- then $X \sim \mathcal{N}(\mu, \sigma^2)$ where

$$\mu = \sum_{i=1}^{n} \mu_i, \quad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$

Example

The yearly precipitation in L.A. is a (independent) normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.

(a) Find the probability that the total precipitation during the next 2 years will exceed 25 inches.

(b) Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches.

Solution

- X_1, X_2: precipitation next 2 years
- $X_1 + X_2 \sim \mathcal{N}(2 \times 12.08, 2 \times (3.1)^2)$
- (a)

$$P(X_1 + X_2 > 25) = P(Z > \frac{25 - 24.16}{\sqrt{19.22}})$$
$$= 1 - \Phi(.1916) = .4240$$

- $-X_2 \sim \mathcal{N}(-12.08, 9.61)$
- $X_1 - X_2 \sim \mathcal{N}(0, 19.22)$

$$\begin{aligned} P(X_1 > X_2 + 3) &= P(X_1 - X_2 > 3) \\ &= P(Z > \frac{3-0}{\sqrt{19.22}}) \\ &= 1 - \Phi(.6834) = .2469 \end{aligned}$$

Homework

Chapter 5: 23, 24, 25, 26, 31