Chapter 2 part 4 §2.2.3 The Normal Distribution

also called "Gaussian distribution", model for measurement errors; person's height; distribution of IQ scores; the velocity of a gas molecule; more generally, $X_1, X_2, \cdots X_n$ i.i.d. $S_n = X_1 + X_2 + \cdots + X_n \sim$ normal,

$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \sim$ normal

definition: $X \sim N(\mu, \sigma^2)$,

p.d.f. $X \sim f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$

"μ and σ" called "mean and standard deviation" of r.v. X.

standard normal r.v. if $\mu = 0, \sigma = 1$. $N(0,1) =: Z$.

$$Z \sim f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty.$$

$$\int_{-\infty}^{\infty} f(z)\,dz = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz = \sqrt{2\pi}.$$

using R, plot some p.d.f.'s with μ and σ.

Calculation: $Z \sim N(0,1)$,

$P(Z \le z) = \Phi(z)$ from z-table.

In R, pnorm(z), or pnorm(z, 0, 1).

Expl. $P(Z \le 1.26) = \text{pnorm}(1.26) = 0.89616.$

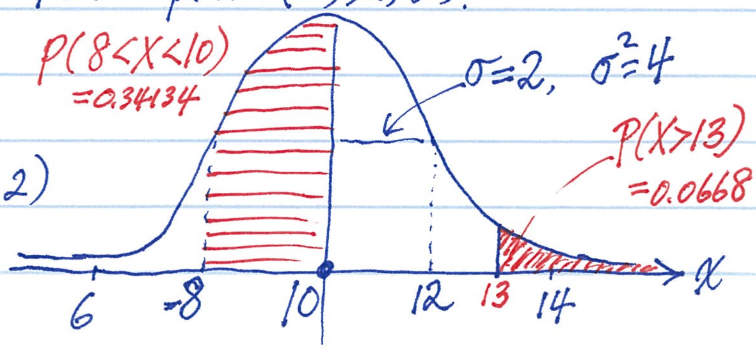
$P(Z < -0.86) = \text{pnorm}(-0.86) = 0.1949$

$$\begin{aligned} P(-1.25 < Z < 0.37) &= P(Z < 0.37) - P(Z < -1.25) \\ &= \text{pnorm}(0.37) - \text{pnorm}(-1.25) \\ &= 0.64431 - 0.10565 \\ &= 0.53866 \end{aligned}$$

For $X \sim N(\mu, \sigma^2)$, $P(X \le x) = F(x) = \text{pnorm}(x, \mu, \sigma)$.

$X \sim N(10, 2^2)$.

$$\begin{aligned} P(X > 13) &= 1 - P(X < 13) \\ &= 1 - \text{pnorm}(13, 10, 2) \\ &= 1 - 0.9332 \\ &= 0.0668 \end{aligned}$$



$$P(8<X<10) = P(X<10) - P(X<8)$$
$$= \text{pnorm}(10, 10, 2) - \text{pnorm}(8, 10, 2)$$
$$= 0.5 - 0.15866$$
$$= 0.34134.$$

The Beta density.

$X \sim Beta(a,b)$, if $X \sim f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1} \cdot (1-u)^{b-1}, & 0 \le u \le 1 \\ 0, & \text{else.} \end{cases}$

Using R to plot pdf's

with (a) $a=2, b=2$,
(b) $a=6, b=2$,
(c) $a=6, b=6$,
(d) $a=0.5, b=4$.

```
x <- seq(0, 1, 0.01)
df1 <- dbeta(x, 2, 2)
df2 <- dbeta(x, 6, 2)
df3 <- dbeta(x, 6, 6)
df4 <- dbeta(x, 0.5, 4)
plot(x, df1, type="l")
plot(x, df2, type="l")
plot(x, df3, type="l")
plot(x, df4, type="l")
```

or

```
plot(x, df1, type="l")
lines(x, df2, lty=2)
.......
```

put 4 curves on the same graph!

or

```
par(mfrow=c(2,2))
plot(x, df1, type="l")
plot(x, df2, type="l")
plot(x, df3, type="l")
plot(x, df4, type="l")
```

put 4 densities on the same page!

§2.3. Functions of a random variable

r.v. $X \sim f(x)$, $Y = g(X)$, what is the p.d.f for r.v. Y?

e.g. X is the velocity of a particle of mass m,
interested in p.d.f of the particle's kinetic energy, $Y = \frac{1}{2} m X^2$.

notation: $X \sim f_X(x)$, $Y = g(X)$, $Y \sim f_Y(y) = ?$

Expl. Let $X \sim U[0,1]$, find p.d.f of $Y = X^2$.

Ans: step 1. find c.d.f of Y, $F_Y(y)$.

step 2. p.d.f $f_Y(y) = F_Y'(y)$.

step 1.

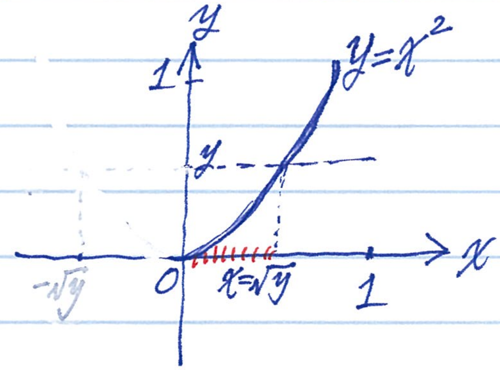
since $X \sim U[0,1]$, X takes value in $[0,1]$. $Y = X^2$ takes value in $[0,1]$ too.

thus, if $y \le 0$. $F_Y(y) = P(Y \le y) = 0$.

if $y \ge 1$. $F_Y(y) = P(Y \le y) = 1$.

if $0 \le y \le 1$, $F_Y(y) = P(Y \le y) = P(X^2 \le y)$

$= P(0 \le X \le \sqrt{y}) = P(0 \le X \le \sqrt{y})$



$= \int_0^{\sqrt{y}} 1\, dx = \sqrt{y}$.

i.e. $F_Y(y) = \begin{cases} 0, & y \le 0 \\ \sqrt{y}, & 0 \le y \le 1 \\ 1, & y \ge 1 \end{cases}$

step 2. $f_Y(y) = \begin{cases} 0, & \text{else} \\ \frac{1}{2\sqrt{y}}, & 0 < y \le 1 \end{cases}$ or $f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}, & 0 < y \le 1 \\ 0, & \text{else.} \end{cases}$

Expl. $X \sim U[0,1]$, find p.d.f of $Y = e^X$.

Ans: $X \sim f_X(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \text{else.} \end{cases}$ $y = e^x$, range $(e^0, e^1) = (1, e)$.

step 1. $1 \le y \le e$

$F_Y(y) = P(Y \le y) = P(e^X \le y) = P(X \cdot \ln e \le \ln y)$

$= P(X \le \ln y) = \int_0^{\ln y} 1 \cdot dx = \ln y$.

$F_Y(y) = \begin{cases} 0, & y \le 1 \\ \ln y, & 1 \le y \le e \\ 1, & y \ge e. \end{cases}$

confirm: $\int_{-\infty}^{\infty} f_Y(y)\, dy$

$= \int_1^e \frac{1}{y}\, dy$

$= \ln y \Big|_1^e$

$= \ln e - \ln 1 = 1$ ✓

step 2. $f_Y(y) = F_Y'(y) = \begin{cases} \frac{1}{y}, & 1 \le y \le e \\ 0, & \text{else.} \end{cases}$

Proposition A: If $X \sim N(\mu, \sigma^2)$, $Y = aX + b$,
then $Y \sim N(a\mu + b, a^2\sigma^2)$.

$a > 0$.

proof: step1. $F_Y(y) = P(Y \le y) = P(aX + b \le y) = P\left(X \le \frac{y-b}{a}\right)$

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

step2. $f_Y(y) = F_Y'(y) = \left[\int_{-\infty}^{\frac{y-b}{a}} f_X(x)\,dx\right]_y'$

$F(u) = \int_a^u f(t)\,dt$

$$= f_X\left(\frac{y-b}{a}\right) \cdot \left(\frac{y-b}{a}\right)'$$

$\Rightarrow F'(u) = f(u)$

$$= \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

$$= \frac{1}{a} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\, a \cdot \sigma} e^{-\frac{(y - a\mu - b)^2}{2a^2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi}\,(a \cdot \sigma)} e^{-\frac{[y - (a\mu + b)]^2}{2(a\cdot\sigma)^2}} \sim N(a\mu + b, a^2\sigma^2).$$

i.e. $Y \sim f_Y(y)$, $Y \sim N(a\mu + b, a^2\sigma^2)$. ($a < 0$ can be shown similarly)

Expl A. Scores on a certain standardized test, IQ scores, are approximately normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 15$. Here we are are referring to the distribution of scores over a very large population, and we approximate that discrete c.d.f by normal c.d.f. An individual is selected at random. What's prob. that his score X satisfies $120 < X < 130$?

Ans: $X \sim N(100, 15^2)$.

$P(120 < X < 130) = P(X < 130) - P(X < 120)$

$= \text{pnorm}(130, 100, 15) - \text{pnorm}(120, 100, 15)$ ← In R

$= 0.9772 - 0.9082 = 0.069.$

Another way is to use Z-table.

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

thus $P(120<X<130) = P\left(\frac{120-100}{15} < \frac{X-100}{15} < \frac{130-100}{15}\right)$

$= P(1.33 < Z < 2) = P(Z<2) - P(Z<1.33)$

$= \Phi(2) - \Phi(1.33) = 0.9772 - 0.9082 = 0.069$

Expl B. $X \sim N(\mu, \sigma^2)$. find prob. that $P(|X-\mu| < \sigma)$.

Ans. $P(|X-\mu|<\sigma) = P\left(\left|\frac{X-\mu}{\sigma}\right| < 1\right) = P(|Z|<1)$

$= P(-1<Z<1)$

$= P(Z<1) - P(Z<-1) = 0.68.$

Expl C. $Z \sim N(0,1)$. find p.d.f. of $X = Z^2$.

Ans. step 1. $F_X(x) = P(X \le x) = P(Z^2 \le x)$. (only need to consider $x \ge 0$. $F_X(x) = 0$, if $x \le 0$)

$= P(-\sqrt{x} \le Z \le \sqrt{x})$

$= P(Z \le \sqrt{x}) - P(Z \le -\sqrt{x})$

$= \int_{-\infty}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - \int_{-\infty}^{-\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$ (use $\left[\int_{-\infty}^{u} f(t)dt\right]'_u = f(u)$)

step 2. $f_X(x) = F_X'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{x})^2}{2}} \cdot (\sqrt{x})'_x - \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{x})^2}{2}} \cdot (-\sqrt{x})'_x$

$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \cdot \frac{1}{2\sqrt{x}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \cdot \left(-\frac{1}{2\sqrt{x}}\right)$

$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\frac{x}{2}} = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-\frac{1}{2}x}$. $x \ge 0$.

thus, $X = Z^2 \sim f_X(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-\frac{1}{2}x}$, $x > 0$.

Note: $X \sim Gamm(\alpha, \lambda)$, $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$. if $\alpha = \lambda = \frac{1}{2}$, $\Gamma(\alpha) = \Gamma(\frac{1}{2}) = \sqrt{\pi}$

we have $f(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{1}{2}x}$. i.e., $X = Z^2 \sim Gamma(\frac{1}{2}, \frac{1}{2})$.

$X = Z^2$, $Z \sim N(0,1)$, X called chi-square r.v. with df = 1.

notation: $Z^2 \sim \chi^2(1)$.

further, Z_1, Z_2, indep, all $N(0,1)$. then $Z_1^2 + Z_2^2 \sim \chi^2(2)$.

Expl. D. $U \sim U[0,1]$, $V = 1/U$, find p.d.f. of V.

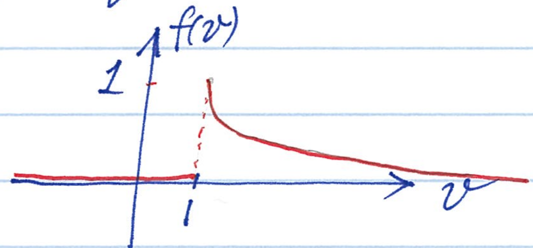
Ans. step 1. $V = 1/U$, $u \in [0,1]$, $V \in (1, \infty)$, $v \geq 1$.

$$F_V(v) = P(V \leq v) = P\left(\frac{1}{U} \leq v\right) = P\left(U \geq \frac{1}{v}\right)$$

$$= \int_{\frac{1}{v}}^{1} 1\, du = 1 - \frac{1}{v}. \quad v \geq 1.$$

step 2. $f_V(v) = F_V'(v) = \left(1 - \frac{1}{v}\right)' = \frac{1}{v^2}$, $v \geq 1$.

$$V \sim f_V(v) = \begin{cases} \frac{1}{v^2}, & v \geq 1 \\ 0, & v < 1 \end{cases}$$

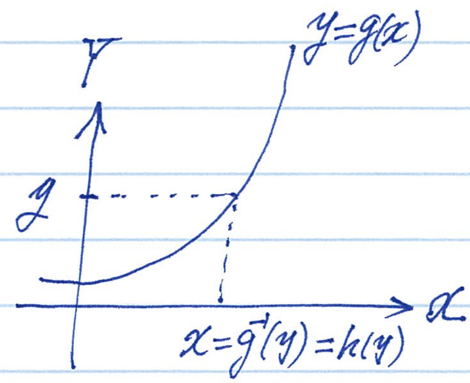


Proposition B. $X \sim f(x)$, $Y = g(X)$, if $g(x)$ is strictly monotonic function, differentiable.

$y = g(x)$. $x = g^{-1}(y) = h(y)$,

then $Y \sim f_Y(y) = f_X(h(y)) \cdot |h'(y)|$

$$= f_X(g^{-1}(y)) \cdot \left|\frac{d}{dy} g^{-1}(y)\right|.$$



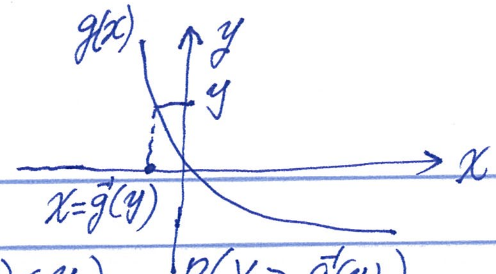
proof: Assume $g(x)$ increasing function in x.

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y))$$

$$= \int_{-\infty}^{g^{-1}(y)} f_X(x)\,dx = \int_{-\infty}^{h(y)} f_X(x)\,dx.$$

$$f_Y(y) = F_Y'(y) = \left[\int_{-\infty}^{h(y)} f_X(x)\,dx\right]_y' = \cancel{f_X(h(y))}\ f_X(h(y)) \cdot h'(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

if $g(x)$ is decreasing,



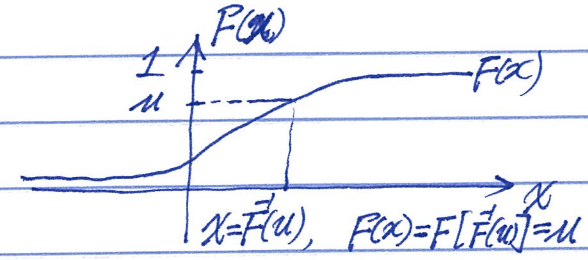
$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \ge g^{-1}(y))$$
$$= 1 - P(X \le g^{-1}(y)) = 1 - P(X \le h(y))$$
$$= 1 - \int_{-\infty}^{h(y)} f_X(x)\,dx$$

$$f_Y(y) = F_Y'(y) = 0 - f_X(h(y)) \cdot h'(y)$$
$$= f_X(h(y))(-h'(y)) = f_X(h(y))\,|h'(y)|, \text{ since } h'(y) < 0.$$

Proposition C. Let $U = F(X)$, where $X \sim f$, cdf is F.
then $U \sim U[0,1]$.

proof: since $u = F(x)$, u's range is over $[0,1]$.

$$F_U(u) = P(U \le u) = P(F(X) \le u) = P(X \le F^{-1}(u))$$
$$= F[F^{-1}(u)] = u. \quad 0 \le u \le 1.$$



$$f_U(u) = F_U'(u) = 1.$$

i.e., $U \sim f_U(u) = \begin{cases} 1, & 0 \le u \le 1, \\ 0, & \text{else.} \end{cases}$ $U \sim U[0,1]$.

Prop. D. $U \sim U[0,1]$, $X = F^{-1}(U)$, then X has c.d.f. F.

proof: $F_X(x) = P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$.

i.e. X has c.d.f. $F(x)$.