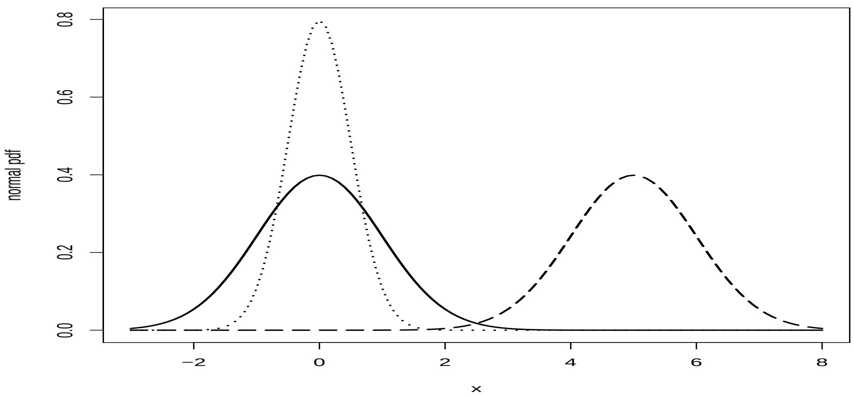
Common univariate distributions

Continuous distributions: Normal (Gaussian) distributon

$X \sim N(\mu, \sigma^2) \quad -\infty < \mu < \infty, \sigma > 0$

- pdf given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$



- Motivation: single most important distribution
  - widely used & analytically tractable
  - bell-shaped density seems to occur naturally
  - Central Limit Theorem (normal distribution is extremely relevant in large samples; more later)

- $\mu \in \mathbb{R}$ is the mean $\mathrm{E}X$ of the distribution

- $\sigma^2 = \mathrm{Var}(X)$ is the variance of the distribution; σ is the standard deviation

- Many properties of the normal distribution can be most easily derived using the $N(0,1)$ or **standard normal distribution**

1. If $X \sim N(\mu, \sigma^2)$ then $Z = (X - \mu)/\sigma \sim N(0,1)$.
2. If $Z \sim N(0,1)$, then $X = a + bZ \sim N(\mu = a, \sigma^2 = b^2)$ for $a, b \in \mathbb{R}$

Common univariate distributions

Standard normal distribution (cont'd)

In developing normal distributions, one actually *starts* with defining the standard normal distribution:

Definition: If a random variable Z has pdf $Z \sim N(0,1)$

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

then we say Z has a **standard normal distribution**, denoted $N(0,1)$.

Notation:

1. it's fairly standard (at least in statistics) to use $\phi(z)$ to denote the standard normal pdf

$$\underset{\text{pdf}}{\phi}(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

2. it's also standard to use $\Phi(z)$ to denote the standard normal cdf

$$\Phi(z) = \mathbb{P}(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy$$

$$\frac{d\Phi(z)}{dz} = \phi(z)$$

3. evaluating Φ is a numerical analysis problem because the pdf ϕ has no simple anti-derivative (the cdf Φ is often tabulated and can be computed with software too)

Common univariate distributions

Standard normal distribution (cont'd)

- Mean: $Z \sim N(0,1)$

$$EZ = \int_{-\infty}^{\infty} \underbrace{z \overbrace{\frac{1}{\sqrt{2\pi}} e^{-z^2/2}}^{\text{even}}}_{\text{odd}} dz = 0$$

- Variance:

$$1 = \text{Var}(Z) = \underline{\underline{EZ^2}} = \int_{-\infty}^{\infty} \underbrace{z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2}}_{\text{even}} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-z^2/2} dz$$

$$\text{Var}(Z) = EZ^2 - (\underbrace{EZ}_{0})^2$$

$$\boxed{\sqrt{y} = \frac{z}{\sqrt{2}}} \qquad y = z^2/2 \qquad dy = \frac{2z}{2} dz \qquad dz = \frac{dy}{z} = \frac{dy}{\sqrt{2y}}$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} (2y)\, e^{-y} \frac{dy}{\sqrt{2y}}$$

$$= \frac{2\,2}{\sqrt{2}\sqrt{2\pi}} \int_0^{\infty} y^{1/2} e^{-y} dy$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\infty} y^{1/2} e^{-y} dy \qquad \Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy$$

$$= \frac{2}{\sqrt{\pi}} \underbrace{\int_0^{\infty} y^{1/2+1-1} e^{-y} dy} = \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) \overset{(**)}{=} \frac{\cancel{2}}{\cancel{\sqrt{\pi}}} \cdot \frac{1}{\cancel{2}} \cancel{\sqrt{\pi}} = 1$$

$(**)$: $\Gamma(z+1) = z\Gamma(z)$

$\Gamma\left(\frac{3}{2}\right) = \Gamma\left(\frac{1}{2}+1\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}$

- mgf:

$$\underbrace{M_Z(t)} \overset{\text{def}}{=} Ee^{tZ} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2tz)} dz$$

$$\overset{(*)}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(z-t)^2 - t^2]} dz$$

$$= e^{t^2/2} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(z-t)^2]} dz}_{=1}$$

$$= e^{t^2/2}, \quad \forall t \in \mathbb{R}.$$

$$M_Z(t) = e^{t^2/2}$$

$(a-b)^2 = a^2 + b^2 - 2ab$

$z^2 - 2tz = z^2 + t^2 - t^2 - 2tz$

$= (z^2 + t^2 - 2tz) - t^2$

$\overset{(*)}{=} (z-t)^2 - t^2$

Note: Let $Y \sim N(t, 1)$

$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-t)^2}$

Common univariate distributions

From standard normal to normal

Step 1: Take any μ + σ^2

Step 2: get a $Z \sim N(0,1)$

Step 3: Construct $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

Definition: If, for $\mu \in \mathbb{R}$ and $\sigma > 0$, X has the same distribution as

$$\mu + \sigma Z$$

for $Z \sim N(0, 1)$ (standard normal), then we say X has a normal $N(\mu, \sigma^2)$ distribution.

Some facts/properties of $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$:

1. cdf: for $-\infty < x < \infty$,

$$F_X(x) = P(X \leq x) = P(\mu + \sigma Z \leq x) = \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right) \overset{(*)}{=} \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Φ: CDF of standard $N(0,1)$

2. pdf: for $-\infty < x < \infty$,

$$f_X(x) = \frac{d}{dx}F_X(x) \overset{(*)}{=} \frac{d\,\Phi\left(\frac{x-\mu}{\sigma}\right)}{dx} = \frac{1}{\sigma}\underbrace{\phi\left(\frac{x-\mu}{\sigma}\right)}$$

ϕ: density

$$\boxed{= \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

$$\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

3. Mean: $\mathrm{E}X = \mathrm{E}(\mu + \sigma Z) = \mu + \cancel{\sigma \mathrm{E}Z}_{0} = \mu$

4. Variance: $\mathrm{Var}(X) = \mathrm{Var}(\mu + \sigma Z) = \sigma^2 \underbrace{\cancel{\mathrm{Var}(Z)}}_{=1} = \sigma^2$

Common univariate distributions

From standard normal to normal (cont'd)

Given $\mu, \sigma^2, Z \Rightarrow X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

5. mgf: for $-\infty < t < \infty$,

$$M_X(t) \underset{\text{def}}{=} \mathrm{E}e^{tX} \underset{X=\mu+\sigma Z}{=} \mathrm{E}e^{t(\mu+\sigma Z)} = \mathrm{E}e^{t\mu}e^{t\sigma Z} = e^{t\mu}\mathrm{E}e^{t\sigma Z} = e^{t\mu} M_Z(t\sigma)$$

$$= e^{t\mu} e^{\frac{t^2\sigma^2}{2}}$$

$$\circledast = e^{\mu t + \frac{t^2\sigma^2}{2}} \Rightarrow M_X(t) \Rightarrow X \sim N(\mu, \sigma^2)$$

6. If $a, b \in \mathbb{R}$ with $b \neq 0$ and $X \sim N(\mu, \sigma^2)$, then $Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$

$\mathbb{E}(Y) = \mathbb{E}(a+bX) = a + b\mathbb{E}X = a + b\mu$

$Var(Y) = Var(a+bX) = b^2\sigma^2$

$Y \sim N(a+b\mu, b^2\sigma^2)$

$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(a+bX)}] = \mathbb{E}[e^{ta} e^{tbX}] = e^{ta}\underbrace{\mathbb{E}[e^{tbX}]}_{\circledast} = e^{ta} M_X(tb)$

$\circledast = e^{ta} e^{\mu tb + \frac{(tb)^2\sigma^2}{2}}$

$= e^{(a+b\mu)t + \frac{t^2}{2} b^2\sigma^2}$

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

$b = \frac{1}{\sigma}$, $a = -\mu/\sigma$

$Z = -\frac{\mu}{\sigma} + \frac{X}{\sigma} \Rightarrow Z \sim N(\underbrace{\cancel{\frac{-\mu}{\sigma}} + \cancel{\frac{1}{\sigma}\mu}}_{=0}, \underbrace{\cancel{\frac{1}{\sigma^2}\sigma^2}}_{=1}) = N(0,1)$

8. 68-95-99% rule (prob that X lies within 1,2, or 3 standard deviations of μ)

$$\begin{aligned}
P(\mu - \sigma \le X \le \mu + \sigma) &= F_X(\mu + \sigma) - F_X(\mu - \sigma) &= \Phi(1) - \Phi(-1) &= 0.6828 \\
P(\mu - 2\sigma \le X \le \mu + 2\sigma) &= F_X(\mu + 2\sigma) - F_X(\mu - 2\sigma) &= \Phi(2) - \Phi(-2) &= 0.9544 \\
P(\mu - 3\sigma \le X \le \mu + 3\sigma) &= F_X(\mu + 3\sigma) - F_X(\mu - 3\sigma) &= \Phi(3) - \Phi(-3) &= 0.9974
\end{aligned}$$

9. To recap:

   (a) If $Z \sim N(0,1)$ and $\mu \in \mathbb{R}$, $\sigma > 0$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

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

   (c) In general, if $a, b \in \mathbb{R}$ with $b \neq 0$ and $X \sim N(\mu, \sigma^2)$, then $a + bX \sim N(a + b\mu, b^2\sigma^2)$

ASIDE

Common univariate distributions

Distributions related to normal

1. Log-normal: If $X \sim N(\mu, \sigma^2)$ then $Y = e^X \sim \text{LogNormal}(\mu, \sigma^2)$
   - pdf given by
   $$f_Y(y) = \frac{1}{y} f_X(\log y) = \frac{1}{y}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2}, \quad 0 < y < \infty$$
   - Mean: $\text{E}Y = \text{E}e^X = M_X(t=1) = e^{\mu + \frac{1}{2}\sigma^2}$

     Note: $\text{E}Y = \text{E}e^X \geq e^{\text{E}X} = e^{\mu}$ by Jensen's inequality
   - shape is like that of a gamma distribution with $\alpha > 1$
   - common in economics (e.g., assume log(income) is normal) and also as a failure time distribution

$F_Y(y) \overset{\text{def}}{=} \mathbb{P}(Y \leq y) = \mathbb{P}(e^X \leq y)$
$= \underbrace{\mathbb{P}(X \leq \ln y)}$
$= F_X(\ln y)$
$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\ln y)$
$= \frac{1}{y} f_X(\ln y)$

2. If $Z \sim N(0,1)$ then $Z^2 \sim \chi^2_1$ (chi-squared 1 df)

3. Standard Browian motion $\{B(t) : t \geq 0\}$
   - a stochastic process where $B(t)$ represents an object's position at time t
   - $B(0) = 0$ and $B(t) \sim N(\mu = 0, \sigma^2 = t)$ for each $t > 0$
   - increments $B(s)$ and $B(t+s) - B(s)$ are independent, any $t > s \geq 0$
   - appears often in probability/statistics/economics