NORMAL APPROXIMATION

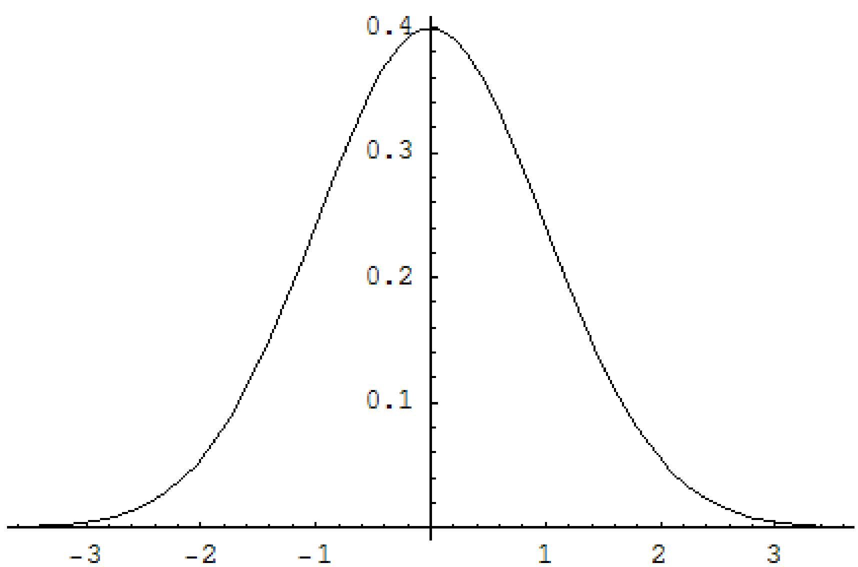
Standardized Normal Distribution

`Standardized` implies that its mean is equal to 0 and the standard deviation is equal to 1. We will always use Z as a name of this RV, $\mathcal{N}(0,1)$ will be our symbolic notation for the corresponding distribution.

In the last chapter we discovered that, when sampling from 'almost' any distribution, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ has a `sampling distribution` whose MGF is $e^{\frac{t^2}{2}}$. We will show that this corresponds to: $f(z) = c \cdot e^{-\frac{z^2}{2}}$, where $-\infty < z < \infty$. c is a constant whose value is a reciprocal of $I \equiv \int\limits_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz$.It is actually easier (an understatement) to compute: $I^2 = \int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx \times \int\limits_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy = \iint\limits_{x-y\text{ plane}} e^{-\frac{x^2+y^2}{2}}\,dx\,dy = \int\limits_0^{2\pi}\int\limits_0^{\infty} e^{-\frac{r^2}{2}} r\,dr\,d\varphi = 2\pi \int\limits_0^{\infty} e^{-u}\,du = 2\pi$. Thus $I = \sqrt{2\pi}$ and $c = \frac{1}{\sqrt{2\pi}}$:

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

(a symmetric 'bell-shaped' curve):



To verify that this is the correct answer: $M(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \cdot e^{zt}\,dz = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{z^2}{2}+zt}\,dz =$

$\frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int\limits_{-\infty}^{\infty} e^{-\frac{(z-t)^2}{2}} dz = \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int\limits_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = e^{\frac{t^2}{2}}$ (check). From the expansion $M(t) = 1 + \frac{t^2}{2} + \frac{\left(\frac{t^2}{2}\right)^2}{2} + ...$ we can immediately establish that $\mu = 0$ (all *odd* moments equal to 0), $\sigma = 1$, and the kurtosis of Z is equal to 3. There is no 'analytic' expression for $F(x)$, but Maple has no difficulty evaluating any probability we need numerically, e.g. $\Pr(-1.3 < Z < 0.5) = \frac{1}{\sqrt{2\pi}} \int_{-1.3}^{0.5} \exp(-\frac{z^2}{2}) dz = 0.5947$

General Normal distribution

We define Z as the following *linear transformation* of Z

$$X = \sigma Z + \mu$$

where $\sigma > 0$ and μ are two constants. From what we know about linear transformations, $\mathbb{E}(X) = \mu$, $\text{Var}(X) = \sigma^2$, and $M_x(t) = e^{\mu t} M_z(\sigma t) = e^{\frac{\sigma^2 t^2}{2} + \mu t}$. The *shape* of the PDF remains the same, only the scale changes.

EXAMPLE: If $M(t) = e^{-2t+t^2}$, what is the distribution? Answer: $\mathcal{N}(-2, \sqrt{2})$.

Note that any further linear transformation of $X \in \mathcal{N}(\mu, \sigma)$, such as $Y = aX + b$, keeps the result Normal. Also: when X_1 and X_2 are *independent* Normal RVs with any (mismatched) parameters, i.e. $X_1 \in \mathcal{N}(\mu_1, \sigma_1)$ and $X_2 \in \mathcal{N}(\mu_2, \sigma_2)$, then their *sum* $X_1 + X_2$ is *also* Normal, which follows from $M_{X_1+X_2}(t) = e^{(\mu_1+\mu_2)t + \frac{\sigma_1^2+\sigma_2^2}{2} t^2}$.

To find $f_x(x)$, we do this: $F_x(x) = \Pr(X < x) = \Pr(\sigma Z + \mu < x) = \Pr(Z < \frac{x-\mu}{\sigma}) = F_z\left(\frac{x-\mu}{\sigma}\right)$. Differentiating with respect to x yields: $f_x(x) = \frac{1}{\sigma} f_z\left(\frac{x-\mu}{\sigma}\right) =$

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad -\infty < x < \infty$$

EXAMPLE: Knowing that $f(x) = \frac{1}{3\sqrt{2\pi}} \exp(-\frac{x^2+4x+4}{18})$, identify the distribution. Answer: $\mathcal{N}(-2, 3)$.

Computing probabilities is easy. EXAMPLE: If $X \in \mathcal{N}(17,3)$, $\Pr(10 < X < 20) = \frac{1}{3\sqrt{2\pi}} \int_{10}^{20} \exp\left(-\frac{(x-17)^2}{2\times 3^2}\right) dx = 0.8315$

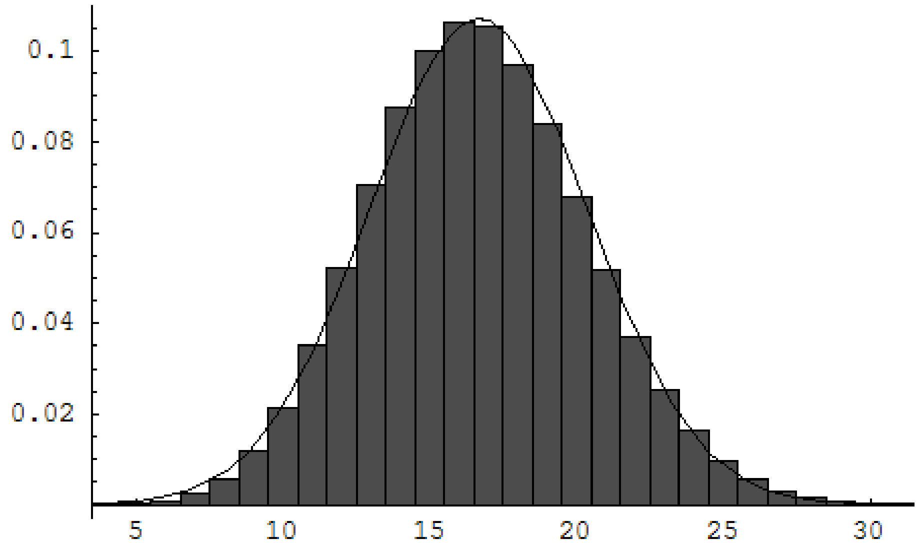
For a Normally distributed RV, the $\mu \pm \sigma$ interval contains 68.26% of the total probability, $\mu \pm 2\sigma$ contains 95.44%, and $\mu \pm 3\sigma$ raises this to a 'near certain' 99.74% (for any practical purpose, the range is 'finite').

Applications of Central Limit Theorem:

Finally, we can apply our knowledge that the distribution of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is, approximately, $\mathcal{N}(0,1)$, the bigger n, the better the approximation. This can be re-stated as: $\bar{X} \tilde{\in} \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$ or, equivalently: $X_1 + X_2 + ... + X_n \tilde{\in} \mathcal{N}(n\mu, \sqrt{n}\sigma)$.

EXAMPLES:

- Roll a die 100 times, what is the probability of getting more than 20 sixes? $\triangleright$ X has the binomial distribution with $n = 100$ and $p = \frac{1}{6}$. Since n is 'large', its distribution will be quite close to $\mathcal{N}(\frac{50}{3}, \sqrt{\frac{125}{9}})$. Thus $\Pr(20 < X_{\text{Binomial}}) \approx \Pr(20.5 < X_{\text{Normal}}) = \frac{1}{\sqrt{125/9\times 2\pi}} \int_{20.5}^{\infty} \exp(-\frac{(x-50/3)^2}{2\times 125/9})dx = 15.18\%$ (the exact answer is 15.19% - in this case, one would expect to be within 0.5% of the exact answer). Note the continuity correction, clear from:



- If X has the Poisson distribution with $\lambda = 35.14$, approximate $\Pr(X \leq 30)$. $\rhd$ Since $X \tilde{\in} \mathcal{N}(35.14, \sqrt{35.14})$, $\Pr(X_{\text{Normal}} < 30.5) = \frac{1}{\sqrt{35.14 \times 2\pi}} \int_{-\infty}^{30.5} \exp(-\frac{(x-35.14)^2}{2 \times 35.14}) dx =$ 21.69% (the exact answer is 22.00%).

- Consider rolling a die repeatedly until obtaining 100 sixes. What is the probability that this will happen in fewer than 700 rolls? $\rhd$ The exact distribution of X is Negative Binomial, with $p = \frac{1}{6}$ and $k = 100$. $\Pr(X_{\text{Normal}} < 699.5) = \frac{1}{\sqrt{3000 \times 2\pi}} \int_{-\infty}^{699.5} \exp(-\frac{(x-600)^2}{2 \times 3000}) dx =$ 96.54% (the exact answer is 96.00%).

- If 5 cards are deal from a standard deck of 52, repeatedly and independently 100 times, what is the probability of dealing at least 50 aces in total? $\rhd$ We need $\Pr(X_1 + X_2 + ... + X_{100} \geq 50)$, where the X_i's are independent, *hypergeometric*, with $N = 52$, $K = 4$ and $n = 5$ ($\mu = \frac{5}{13}$ and $\sigma = \sqrt{\frac{5}{13} \times \frac{12}{13} \times \frac{47}{51}}$ *each*). For their sum, $\mu_{\text{sum}} = \frac{500}{13}$ and $\sigma_{\text{sum}} = \sqrt{\frac{500}{13} \times \frac{12}{13} \times \frac{47}{51}} = \sqrt{\frac{94000}{2873}}$. The answer is, approximately, $\frac{1}{\sqrt{94000/2873 \times 2\pi}} \int_{49.5}^{\infty} \exp(-\frac{(s-500/13)^2}{2 \times 94000/2873}) ds = 2.68\%$ (the exact answer is 3.00%).

- Consider a random independent sample of size 200 form the uniform distribution $\mathcal{U}(0,1)$. Find $\Pr(0.49 \le \bar{X} \le 0.51)$. $\triangleright$ We know that $\bar{X} \tilde{\in} \mathcal{N}\left(0.5, \sqrt{\frac{1}{12\times 200}}\right)$, which implies that the above probability is, approximately, $\frac{1}{\sqrt{2\pi/2400}}\int_{0.49}^{0.51}\exp(-\frac{(\bar{x}-.5)^2}{2/2400})d\bar{x} =$ 37.58%.(the exact answer is 37.56% - the approximation is now a lot more accurate for two reasons: the uniform distribution is continuous and symmetric).

- Consider a random independent sample of size 100 from

$X =$	-1	0	1	2
Prob:	$\frac{3}{6}$	$\frac{2}{6}$	0	$\frac{1}{6}$

  What is the probability that the sample total will be negative (losing money, if this represents a game)? $\triangleright$ First we compute the distribution's $\mu = -\frac{1}{6}$ and $\mathrm{Var}(X) = \frac{3+4}{6} - \frac{1}{36} = \frac{41}{36}$, then we introduce $S = \sum_{i=1}^{100} X_i$ and find $\mu_s = -\frac{100}{6}$ and $\sigma_s = \sqrt{\frac{4100}{36}}$. Answer: $\Pr(S < 0) \approx \frac{1}{\sqrt{4100/36\times 2\pi}}\int_{-\infty}^{-0.5}\exp(-\frac{(s+100/6)^2}{2\times 4100/36})ds = 93.51\%$. (the exact value is 93.21%).

- Pay \$10 to play the following game: 5 cards are dealt from a standard deck, and you receive \$10 for each ace and \$5 for each king, queen and jack. First, find the expected value and standard deviation of your net win. $\triangleright$ $W = 10X + 5Y - 10$ (where X is the number of aces dealt, Y correspondingly counts the total of kings, queens and jacks). This implies that $\mu_w = 10 \times \frac{5}{13} + 5 \times \frac{5\times 3}{13} - 10 = -\frac{5}{13}$ dollars, $\mathrm{Var}(W) = 5(10^2 \cdot \frac{1}{13} \cdot \frac{12}{13} + 5^2 \cdot \frac{3}{13} \cdot \frac{10}{13} - 2 \cdot 10 \cdot 5 \cdot \frac{1}{13} \cdot \frac{3}{13})\frac{47}{51} = 44.988$ and $\sigma_w = \sqrt{44.988} = \$6.7073$.

  Secondly, compute the (approximate) probability of losing more that \$50 after 170 rounds of this game. $\triangleright$ Defining $S \equiv \sum_{i=1}^{170} W_i$, we get $\mu_s = -65.385$ and $\sigma_s = 6.7073 \times \sqrt{170} = 87.452$. Thus, $\Pr(S < -50) \approx \Pr(S_{\text{Normal}} < -52.5) = \frac{1}{87.452\sqrt{2\pi}}\int_{-\infty}^{-52.5}\exp(-\frac{(t+65.385)^2}{2\times 87.452^2})dt =$ 55.86% (the exact answer is 56.12% - not bad considering that the individual probabilities are still well over 2%). Note the unusual continuity correction!