Gaussian distribution

- $X$ has a Gaussian (normal) distribution: $X \sim N(\mu, \sigma^2)$, if

\[ f_X(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R} \]

$\mu$: mean; $\sigma^2$: variance

- Standardized version of $X$, $\tilde{X} = \frac{X - \mu}{\sigma}$ is distributed as $N(0,1)$.

\[ F_{\tilde{X}}(a) = P(\tilde{X} \leq a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) \, du = \Phi(a) \]

\[ F_{\tilde{X}}^c(a) = P(\tilde{X} > a) = \int_{a}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) \, du = 1 - \Phi(a) = Q(a) \]

Notice that $Q(a) = \Phi(-a)$ because of symmetry.

We can write:

\[ P(a < X < b) = P(X < b) - P(X < a) \]

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

\[ = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \]

\[ P(X > b) = P\left(\frac{X - \mu}{\sigma} > \frac{b - \mu}{\sigma}\right) = Q\left(\frac{b - \mu}{\sigma}\right) = \Phi\left(-\frac{b - \mu}{\sigma}\right) \]

Today:

1. The Central Limit Theorem and approximation

2. ML parameter estimate for continuous-type variables

3. The Central Limit Theorem and approximation

Loosely speaking, if many independent random variables are added together, and if each of them are small in magnitude compared to sum, the the sum
has approximately normal distribution.

1. If \(X_i \sim \text{Ber}(p)\) and independent of each other, then \(S_n = X_1 + \ldots + X_n\) is 
\(\text{Bi}(n,p)\). We can approximate \(S_n\) by a gaussian random variable.

\[
P(S_n \leq k) \approx P(W \leq k)
\]

where \(W \sim N(\mu, \sigma^2)\), \(\mu = n \cdot p\), \(\sigma^2 = n \cdot p \cdot (1-p)\). Approximation is mostly accurate if

(i) \(\mu = np\) is moderately large

(ii) \(\sigma = np(1-p)\) is moderately large

(iii) The probability of the event of interest is not too small.

If we plot the cdf of \(S_n\) and \(W\): \(n=10, \ p=0.2, \ np=2, \ \sigma = \sqrt{np(1-p)} = 1.6\)

Observations:

(i) cdf of \(S_n\) is picewise constant, cdf of \(W\) is continuous. They cannot be close everywhere.

(ii) Surprisingly, they are close at midle points:

\[
P(S_n \leq 2) = P(S_n \leq 2.5) \approx P(W \leq 2.5)
\]

* Gaussian approximation with the continuity correction.*

\[
P(S_n \leq k) \approx P(W \leq k + 0.5)
\]

\[
P(S_n > k) = 1 - P(S_n \leq k) \approx 1 - P(W \leq k + 0.5) = P(W > k + 0.5)
\]

\[
P(S_n > k) \approx P(W > k - 0.5)
\]

where \(S_n \sim \text{Bi}(n,p)\) and \(W \sim N(\mu, \sigma^2)\), \(\mu = np\), \(\sigma = \sqrt{np(1-p)}\). In particular,

\[
P(S_n = k) = P(S_n \leq k) - P(S_n \leq k-1)
\]

\[
\approx P(W \leq k + 0.5) - P(W \leq k - 0.5) = \int_{k-0.5}^{k+0.5} f_W(u) \, du = \Phi\left(\frac{k+0.5-m}{\sigma}\right) - \Phi\left(\frac{k-0.5-m}{\sigma}\right)
\]
Theorem (DeMoivre–Laplace limit theorem)

Suppose that $S_n, p \approx Bi(n, p), p \in (0,1)$ fixed.

$$
\lim_{n \to \infty} P \left\{ \frac{S_n - np}{\sqrt{np(1-p)}} \leq c \right\} = \Phi(c)
$$

i.e., for large $n$, $S_n$ has the same cdf as a Gaussian random variable.

What happened to continuity correction?

$$
P \left\{ \frac{S_n - np}{\sqrt{np(1-p)}} \leq c \right\} \approx P \left\{ \frac{X - 0.5 - np}{\sqrt{np(1-p)}} \leq c \right\} \quad \text{with continuity correction}
$$

$$
\approx P \left\{ \frac{X - np}{\sqrt{np(1-p)}} \leq c \right\} \quad \text{n is large so that np(1-p) is large}
$$

3.19. [Heads minus tails]

Suppose a fair coin is flipped 100 times, and $A$ is the event:

$A = \{ \text{number of heads} - \text{number of tails} \geq 10 \}$.

(a) Let $S$ denote the number of heads. Express $A$ in terms of $S$. Specifically, identify which values of $S$ make $A$ true.

(b) Using the Gaussian approximation with the continuity correction, express the approximate value of $P(A)$ in terms of the Q function.

Solution:

(a) Notice that $\# \text{ of tails} = 100 - S$. Hence

$$
A = \{ \text{number of heads} - \text{number of tails} \geq 10 \} = \{ S - (100 - S) \geq 10 \}
$$

$$
= \{ S - 50 \geq 5 \}
$$

$$
= \{ S \geq 55 \} \cup \{ S < 45 \}
$$

(b) $P(\{ S \geq 55 \} \cup \{ S < 45 \}) = P(S \geq 55) + P(S \leq 45)$.

Notice that $P(S \leq 45) \approx P(W \leq 45 + 0.5)$ and $P(S \geq 55) \approx P(W \geq 55 - 0.5)$

where $W \sim N(\mu, \sigma^2)$, $\mu = np = 50$, $\sigma = \sqrt{np(1-p)} = 5$. Hence,
\[ P(\{S \geq 55\} \cup \{S \leq 45\}) = P(W \leq 45.5) + P(W \geq 54.5) \]

\[ = \Phi \left( \frac{45.5 - 5}{5} \right) + \Phi \left( \frac{54.5 - 5}{5} \right) \]

The ML parameter estimate for continuous-type variables.

Suppose that the random variable \( X \) is continuous type and its pdf belongs to a family of distributions \( f_\theta(x) \) (example: \( X \) is exponential random variable, \( \theta \) is the parameter of pdf, i.e., \( X \sim \exp(\theta) \)).

Suppose that we observe \( X = x \). Notice that \( P(X = x) = 0 \) since \( X \) is continuous type. Recall that for \( \varepsilon > 0 \) small enough,

\[ P(\varepsilon < X < x + \varepsilon) \approx 2\varepsilon f_\theta(x) \]

Maximum likelihood estimation for continuous type suggests maximizing \( f_\theta(x) \) over \( \theta \), given the observation is \( X = x \).

\[ \hat{\theta}_{ML} = \arg\max_{\theta > 0} f_\theta(x) \]

3.23. [ML parameter estimation of the rate of a Poisson process]

Calls arrive in a call center according to a Poisson process with arrival rate \( \lambda \) (calls/minute). Derive the maximum likelihood estimate \( \hat{\lambda}_{ML} \) if \( k \) calls are received in a \( t \) minute interval.

We consider two different interpretations of the problem.

(a) \( k \)th call is received at time \( t \).

Notice that the time of \( k \)th arrival is \( T_k \), which has Erlang distribution with parameters \( (k, \lambda) \).

\[ \hat{\lambda}_{ML} = \arg\max_{\lambda > 0} f_{T_k}(t) = \arg\max_{\hat{\lambda} > 0} \frac{\lambda^k t^k e^{-\lambda t}}{(k-1)!} \]

To solve above, we need to solve

\[ \frac{d}{d\lambda} f_{T_k}(t) \bigg|_{\lambda = \hat{\lambda}_{ML}} = 0 \quad \Rightarrow \quad \frac{d}{d\lambda} \left( \lambda^k t^k e^{-\lambda t} \right) \bigg|_{\lambda = \hat{\lambda}_{ML}} = 0. \]

(b) In the first \( t \) minutes we received \( k \) calls, i.e., \( N_t = k \).

Notice that \( N_t \sim \text{Poi}(\lambda t) \)

\[ \hat{\lambda}_{ML} = \arg\max_{\lambda > 0} P(N_t = k) = \arg\max_{\lambda > 0} \frac{e^{-\lambda t} (\lambda t)^k}{k!} \]
\[ \hat{\lambda}_{ML} = \arg \max_{\lambda > 0} P(N_t = k) = \arg \max_{\lambda > 0} \frac{e^{-\lambda t} (\lambda t)^k}{k!} \]

Hence, we need to solve

\[ \frac{d}{d\lambda} P(N_t = k) \bigg|_{\lambda = \hat{\lambda}_{ML}} = 0 \Rightarrow \frac{d}{d\lambda} (e^{-\lambda t} \lambda^k) \bigg|_{\lambda = \hat{\lambda}_{ML}} = 0 \]

Interestingly, they both give the same \( \lambda \). Can you interpret why?