Application of CDF (not required)

Quantile - quantile plot

Q-Q Plot - a tool to check how good your model is.

Example Consider a dataset containing $N$ data points. The histogram (empirical PDF) and empirical CDF is as follows:

Is it a Gaussian distribution?
QQ-Plot

$x_1, x_2, \ldots, x_N$ → sorting → $x[1], x[2], \ldots, x[N]$

partition into $N$ equally spaced intervals
inverse CDF mapping
to generate $z_1, \ldots, z_N$

plot $x[1]$ vs $z_1$
plot $x[2]$ vs $z_2$
\vdots
plot $x[N]$ vs $z_N$
QQ-Plot

Why does it work?

Assume $x_1, \ldots, x_N$ are samples of a random variable $X$.

Hypothesis: These data points are generated from certain random variable $\hat{X}$. Let $F_{\hat{X}}$ be its CDF.

Consider $y_1, \ldots, y_N$ are the equally spaced points of $F_{\hat{X}}$. Then the $z_i$'s are

$$z_i = F_{\hat{X}}^{-1}(y_i).$$

Testing: If $X = \hat{X}$, then for large $N$, we must have

$$z_i = F_{\hat{X}}^{-1}(y_i) \approx x_i.$$

Therefore, we should have a linear function if we plot $x_i$ against $z_i$. 
QQ-Plot

Figure: Left: Poor fit. In fact, the empirical data is generated from a $t$-distribution. Right: Good fit.
4. Gaussian Distribution
Gaussian Distribution

Definition (Gaussian Distribution)

Let $X$ be an **Gaussian random variable**. The PDF of $X$ is

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

(11)

where $(\mu, \sigma^2)$ are parameters of the distribution. We write

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

to say that $X$ is drawn from a Gaussian distribution of parameter $(\mu, \sigma^2)$. 

Gaussian Distribution

Figure: Gaussian distribution

Proposition (Mean/Variance of Gaussian Distribution)

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

\[\mathbb{E}[X] = \mu, \quad \text{and} \quad \text{Var}[X] = \sigma^2.\]
Gaussian Distribution

Proof.

\[ E[X] = \int_{-\infty}^{\infty} x f(x) \, dx \]

\[ = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]

let \( u = \frac{x-\mu}{\sigma} \) \( \Rightarrow \frac{du}{\sigma} = dx \)

\[ = \int_{-\infty}^{\infty} (\sigma u + \mu) \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \sigma \, du \]

\[ = \int_{-\infty}^{\infty} \sigma u \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \, du + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \, du \]

\[ = \mu \cdot 0 \]

\[ = \mu \]
Percentile of Gaussian Distribution
Standard Gaussian

Definition (Standard Gaussian)

A **standard Gaussian** (or standard Normal) random variable $X$ has a PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$  \hspace{1cm} (12)

That is, $X \sim \mathcal{N}(0,1)$ is a Gaussian with $\mu = 0$ and $\sigma^2 = 1$.

Definition (CDF of Standard Gaussian)

The $\Phi(\cdot)$ **function** of the standard Gaussian is

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} \, dx$$  \hspace{1cm} (13)
Standardize Random Variable

If \( X \sim \mathcal{N}(\mu, \sigma^2) \), then

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

**Proof.** Key: Change of variable.

\[
\begin{align*}
F_X(x) &= \int_{-\infty}^{x} f_X(x') \, dx' \\
&= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{(x' - \mu)^2}{2\sigma^2}} \, dx' \\
&= \int_{-\infty}^{\frac{x - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x'^2}{2}} \, dx' \\
&= \Phi\left(\frac{x - \mu}{\sigma}\right).
\end{align*}
\]
\[ Z = \frac{X - \mu}{\sigma} \]

\[ E|Z| = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} (E(X) - \mu) \]

\[ = 0 \]

\[ \text{Var}(Z) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} \text{Var}(X - \mu) \]

\[ = \frac{1}{\sigma^2} \text{Var}(X) = \frac{1}{\sigma^2} \sigma^2 = 1. \]
\[ F_z(z) = \int_{-\infty}^{z} f(x) \, dx \]

\[ = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]

Let \( u = \frac{x-\mu}{\sigma} \)

\[ du = \frac{1}{\sigma} \, dx \]

If \( x = z \), then

\[ u = \frac{z-\mu}{\sigma} \]

\[ = \text{CDF of } N(0, 1) \text{ evaluated at } \frac{z-\mu}{\sigma} \]

\[ = \Phi \left( \frac{z-\mu}{\sigma} \right) .\]
Standard Gaussian

\[ \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{x^2}{2}} \, dx \]

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

**Figure**: Definition of \( \Phi(y) \).

**Example.** Let \( X \sim \mathcal{N}(\mu, \sigma^2) \). Find \( \Pr[X \leq b] \) and \( \Pr[a \leq X \leq b] \).

\[ \Pr(X \leq b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \]

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

Example. \( X \sim \mathcal{N}(5, 16) \), find

(a) \( P[X > 3] = 1 - \frac{P(X \leq 3)}{\Phi \left( \frac{3 - \mu}{\sigma} \right)} = \Phi \left( \frac{3 - 5}{4} \right) = \Phi \left( \frac{-1}{2} \right) = 0.3085 \)

(b) If \( P[X < a] = 0.7910 \), find \( a \).

\[
\frac{a - \mu}{\sigma} = 0.81 \quad \Rightarrow \quad a = (0.81)(\sigma) + \mu
\]

(c) If \( P[X > b] = 0.1635 \), find \( b \).

\[
= 1 - P(X \leq b)
\]