quantile probability

QQ-plot and PP-plot

X is a continuous random variable with pdf $f(x)$ and

cdf $F(x) = P(X < x) = \int_{-\infty}^{x} f(x)\,dx$

p-quantile, x_p: Value of x such that $F(x) = p$

(so $0 \leq p \leq 1$) $\Rightarrow$ $x_p = F^{-1}(p)$

X is a r.v. so $F(X)$ is also a r.v.

What is the distribution of $F(X)$?

The cdf of $F(X)$ is $P(F(X) < x) = P(X < F^{-1}(x))$

$= F(F^{-1}(x)) = x$ for $0 \leq x \leq 1$

$\Rightarrow$ $F(X) \sim$ Uniform on $[0, 1]$

Glivenko-Cantelli theorem: The empirical distribution of a sample approaches the theoretical distribution as the sample size tends to ∞.

To check graphically if a sample $\{x_i\}_{i=1}^{n}$ comes from the distribution F:

Order the sample: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

If $\underset{\text{the assumption}}{\overset{}{}}$$X \sim F$ holds then $F(X) \sim U[0,1]$ and

$i/n \approx F(x_{(i)})$ for $i=1,\dots,n$ if n is large

(Explanatory example: Sample 1000 observations from $U[0,1]$. Look at the 250th smallest observation. We expect it to approx. $\frac{250}{1000}$.)

$$i/n \approx F(x_{(i)}) \iff F^{-1}(i/n) \approx x_{(i)}$$

pp-plot: plot the pairs $\left(i/n,\ F(x_{(i)})\right)_{i=1}^{n}$
probabilities

and see if observations scatter along the line $y = x$.

QQ-plot: plot instead the pairs $\left(x_{(i)},\ F^{-1}(i/n)\right)_{i=1}^{n}$
quantile