Fundamental Statistical Measures

Refs: Davis - notes chap. 2

Statistic for a single variable (Moments)

Def: \( X \) a random variable whose values are defined by a random process, a process which produces values not perfectly known (e.g. the Lorenz attractor example)

Many realizations of \( X \) produce an ensemble \( x_n \)

The "expected value" or "mean" or "average"

\[
\langle x \rangle = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} x_n
\]

The expected value has certain properties:

\[
\langle x + y \rangle = \langle x \rangle + \langle y \rangle
\]

\[
\langle aX \rangle = \langle Xa \rangle = a \langle x \rangle \text{ if } a \text{ constant}
\]

\[
\langle xy \rangle \neq \langle x \rangle \langle y \rangle
\]
A complete description of $x$ is given by the distribution function $D_x(X)$

$$D_x(X) = \text{the fraction of } x < X$$

- $X_{\text{Median}}$ is the value of $x$ for which 50% of the $x$ are smaller than $X_{\text{Median}}$ and the remaining 50% is larger.

  $$D_x(+\infty) = 1$$
  $$D_x(-\infty) = 0$$

"The Probability Distribution Function" (PDF)

what is the fraction of occurrences $x$ falling in the interval $X - \frac{dx}{2} < x < X + \frac{dx}{2}$

$$= D_x - D_x - D_x = D_x - D_x$$

$$= (D_x + \frac{dx}{2}) - (D_x - \frac{dx}{2}) = dx$$
The PDF is the probability that any given

The probability that an \( x \) will fall in that interval is given by the slope of the distribution function \( d_{\nu}(x) \). For \( dx \to 0 \)

The slope is given by:

\[
\lim_{dx \to 0} \frac{(d_{\nu} + \frac{dd_{\nu}}{dx}) - (d_{\nu} - \frac{dd_{\nu}}{dx})}{(x + \frac{dx}{2}) - (x - \frac{dx}{2})} = \left( \frac{d}{dx} d_{\nu}(x) \right) = f_{\nu}(x)
\]

\( f_{\nu}(x) \) is "Probability Distribution Function", it is the probability that a given \( x = X \) (the value \( x \))

**NOTE:**

1) \( f_{\nu}(x) \) \( dx = \frac{dd_{\nu}}{dx} dx = dd_{\nu} \)

Multiplying the pdf by \( dx \) gives you the fraction of occurrences of \( x \) falling in the interval \( dx = x_1 - x_2 \)

2) The integral of

\[
\int_{-\infty}^{\infty} f_{\nu}(x) \, dx = \int_{-\infty}^{\infty} \frac{dd_{\nu}}{dx} \, dx = \left[ d_{\nu} \right]_{-\infty}^{+\infty} = d_{\nu}(+\infty) - d_{\nu}(-\infty)
\]

\[
= 1
\]
Example:

Assume the following distribution function.

This is called a Heavyside function. It means that all the values occur when \( x = A \).

\[
P_x(x) = \delta(x - A)
\]

\[
\begin{cases}
0 & \text{for } x \neq A \\
1 & \text{for } x = A
\end{cases}
\]

Recall

\[
P_x(x) = \frac{dD_x}{dx} \quad \Rightarrow \quad D_x(x) = \int_{-\infty}^{x} P_x(x') dx'
\]

\[
= \int_{-\infty}^{x} \delta(x' - A) dx'
\]

Heavyside function.

**HANDY NOTATION FOR PDF**

1. Assume a large number \( N \) of realizations of \( x \).
2. Assume that each \( x \) is one variable \( x_1, x_2, x_3, \ldots, x_N \).

\[
P_x(x) = \frac{1}{N} \left[ \delta(x - x_1) + \delta(x - x_2) + \ldots + \delta(x - x_N) \right]
\]

In the limit \( N \to \infty \)

\[
P(x) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i) = \langle \delta(x - x) \rangle_{\text{expected value}}
\]
It is now instructive to compute the

**moments of** \( x \)

\[
\int_{-\infty}^{\infty} P_x(x) \, dx = \int_{-\infty}^{\infty} <\delta(x-x)> \, dx = <x> = 1
\]

is one when

\( x = x \)

and 0 otherwise

**1st moment**

\[
\mu_1 = \int_{-\infty}^{\infty} x \, P_x(x) \, dx = <x> = \text{expected value or average of} \ x
\]

**2nd moment**

\[
\mu_2 = \int_{-\infty}^{\infty} x^2 \, P_x(x) \, dx = <x^2>
\]

**\( n \)th moment**

\[
\mu_n = \int_{-\infty}^{\infty} x^n \, P_x(x) \, dx = <x^n>
\]

for \( n > 1 \) usually the random variable \( x' \) is defined as an anomaly with respect to the expected value

\( x' = x - <x> \)

\( \mu_2 = \text{variance} = \text{energy of the process} \)

\( \sigma = \sqrt{\mu_2} = \text{standard deviation} = \text{typical fluctuation} \)

Knowledge of all the moments is sufficient to characterize the entire PDF.
Any function of a random variable is also a random variable and its statistics can be determined by the original variable.

Suppose \( y = g(x) = \int_{-\infty}^{+\infty} g(x) \delta(x-x) \, dx \)

If we take the expected value

\[
\langle y \rangle = \langle g(x) \rangle = \int_{-\infty}^{\infty} g(x) \langle \delta(x-x) \rangle \, dx = \int_{-\infty}^{\infty} g(x) p_x(x) \, dx
\]

\[
= \int_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \delta(x-x) \, dx \rightarrow \int_{-\infty}^{\infty} \delta(x-y) \, dy = \int_{-\infty}^{\infty} y p_y(y) \, dy
\]

Now \( dY = dg(x) = \frac{dG}{dx} \, dx \)

\[
\int_{-\infty}^{\infty} Y p_y(Y) \, dY = \int_{-\infty}^{\infty} Y p_y(Y) \frac{dG}{dx} \, dX = \int_{-\infty}^{\infty} g(x) p_x(x) \, dx
\]

Equivalent statements:

\[
P_y(Y) \frac{dG}{dx} = p_x(x)
\]

\[
P_y(Y) \frac{dY}{dx} = p_x(x)
\]

\( \rightarrow \) change of variable!
Aver alternative derivation.

\[ y = f(x) \]

The probability of \( x \) falling between \( x - dx/2 \leq x \leq x + dx/2 \) is equivalent to the probability of \( g(x) \) falling in the interval \( g(x - dx/2) < g(x) < g(x + dx/2) \).

\[ P_x(x) \, dx = P_y(Y) \, dY \]
Example 1:

Assume \( y = 2x \) and \( P(X) = \frac{1}{\sqrt{2\pi\mu_x}} e^{-\frac{x^2}{2\mu_x^2}} \)

Find \( P_y(Y) \)

\[
P_y(Y) \, dY = P_x \, dX
\]

\[
P_y = P_x \left[ \frac{dY}{dX} \right]^{-1}
\]

\[
\frac{dY}{dX} = 2,
\]

\[
P_y = \frac{1}{2} \frac{1}{\sqrt{2\pi\mu_x}} e^{-\frac{(Y/2)^2}{2\mu_x^2}} = \frac{1}{\sqrt{2\pi \cdot 4\mu_x^2}} e^{-\frac{Y^2}{2 \cdot 4\mu_x^2}}
\]

The same Gaussian distribution with 4 \( x \) larger variance or \( d \) with double standard deviation.
Example 2:

\[ x = \ln y \quad \text{with } x \text{ gaussian} \]

\[ p_y = p_x \frac{dx}{dy} = \frac{1}{y} \quad p_x = \frac{1}{\sqrt{2\pi \mu_x}} e^{-\frac{(\ln y)^2}{2\mu_x^2}} \]

\[ \text{log normal distribution} \]
Other PDFs

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

$\mathbb{E}[x] = \text{1st moment} = \int_{-\infty}^{\infty} xp_x \, dx = \mu$

$\text{Var}[x] = \text{2nd moment about } x' = x - \mu = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p_x(x-\mu) \, dx$

Lognormal, Gamma, Exponential, Chi-square, Pearson III, Beta, Gumbel, Weibull, Mixed Exponential...