Moment-generating function

Moment-generating function: definition

Moment-generating function of r.v. X is

$$M_X(t) = \mathbb{E}\left[e^{tX}\right]$$

It does not always exist. If it exists and is finite:

- It uniquely defines distribution of X
- $M_X(t) > 0, \forall t$ and $M_X(0) = 1$
- $M_{aX+b}(t) = e^{bt} M_X(at)$
- For all k exists a finite moment of X and is defined as $\mathbb{E}[X^k] = M_X^{(k)}(0)$ meaning k-th derivative

The purpose of MGF is to replace computation of expectation with differentiation.

Example 1: Bernoulli MGF

Consider $X \sim Be(p)$. What is $M_X(t)$? Find expectation and variance using MGF.

Solution 1

MGF:

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = e^{t\cdot 0} \cdot \mathbb{P}(X=0) + e^{t\cdot 1} \cdot \mathbb{P}(X=1) = q + pe^t$$

First and second derivatives are pe^t, so

$$\mathbb{E}X = M_X'(0) = pe^0 = p = M_X''(0) = \mathbb{E}\left[X^2\right]$$

$$\mathbb{V}\text{ar}(X) = M_X''(0) - \left(M_X'(0)\right)^2 = p - p^2 = p(1-p)$$

Example 2: Poisson MGF

Consider $X \sim Pois(\lambda)$. What is $M_X(t)$? Find expectation and variance using MGF.

Solution 2

MGF:

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \sum_{k=-\infty}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=-\infty}^{\infty} \frac{1}{k!} \left(\lambda e^t\right)^k = \exp\left(\lambda \left(e^t - 1\right)\right)$$

First derivative:

$$M_X'(t) = \lambda e^t \exp\left(\lambda \left(e^t - 1\right)\right)$$

Expectation $M_X'(0) = \lambda$. Second derivative:

$$M_X''(0) = \lambda e^t \exp\left(\lambda \left(e^t - 1\right)\right) + \lambda e^t \exp \lambda e^t \left(\lambda \left(e^t - 1\right)\right)$$

Second moment $M_X''(0) = \lambda + \lambda^2$. Variance $\mathbb{V}\mathrm{ar}(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$.

Example 3: Gaussian MGF

Consider $X \sim \mathcal{N}(\mu, \sigma^2)$. What is $M_X(t)$? Find expectation and variance using MGF.

Solution 3

First let's find for $Y \sim \mathcal{N}(0,1)$, then apply properties.

$$M_Y(t) = \mathbb{E}\left[e^{tY}\right] = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp\left(-\frac{x^2 - 2tx}{2}\right) dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp\left(-\frac{(x-t)^2 - t^2}{2}\right) dx =$$

$$= \exp\left(\frac{t^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp\left(-\frac{(x-t)^2}{2}\right) dx = \exp\left(\frac{t^2}{2}\right)$$

Solution 3 (continued)

From properties, $M_X(t) = e^{\mu t} M_Y(\sigma t) = \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$. First derivative:

$$M_X'(t) = \left(\mu + t\sigma^2\right) \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$$

Second derivative:

$$M''_X(t) = \sigma^2 \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right) + \left(\mu + t\sigma^2\right)^2 \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$$

Expectation: $M'_X(0) = \mu$, variance $M''_X(0) - \left(M'_X(0)\right)^2 = \sigma^2$.

Random vector

Random vector: definition

Consider probability space $(S, \mathbb{P})$. Then, a **random vector** is a function

$$\mathbf{X} : \Omega \to \mathbb{R}^n,$$

where $\mathbf{X} = (X_1, \ldots, X_n)^\top$. Every component X_i of the vector is a random variable. The converse is also true: for any r.v.s $X_1, \ldots, X_n$ a vector $(X_1, \ldots, X_n)^\top$ is a random vector.

Random vector: distribution

The distribution of a random vector $\mathbf{X} = (X_1, \ldots, X_n)^\top$ can be described via **multivariate (joint) cumulative distribution function**:

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n)$$

Properties of multivariate CDF:

- $\lim_{x_i \to -\infty} F_{\mathbf{X}}(\mathbf{x}) = 0$ but $\lim_{x_1, \ldots, x_n \to \infty} F_{\mathbf{X}}(\mathbf{x}) = 1$
- $\lim_{x_i \to \infty} F_{\mathbf{X}}(\mathbf{x}) =$ the function F of everything except x_i
- $F_{\mathbf{X}}(\mathbf{x})$ is non-decreasing and left-continuous in every component
- Supermodulatiry: $F_{\mathbf{X}}(x_1, \ldots, x_i, \ldots, x_n) - F_{\mathbf{X}}(x_1, \ldots, x_i - \varepsilon, \ldots, x_n) \geqslant 0$

Random vector: distribution

If X has continuous distribution, then exists **multivariate (joint) probability density function**, i.e. non-negative function $f_{\mathbf{X}}(\cdot)$ such that

$$\mathbb{P}(\mathbf{X} \in B) = \int_B f_{\mathbf{X}}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

PDF can also be found from CDF:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \dots \partial x_n}$$

Random vector: independence

If all r.v.s X_i are independent, then

$$\begin{cases} F_{\mathbf{X}}(\mathbf{x}) & = \prod_{i=1}^{n} F_{X_i}(x_i), \\ f_{\mathbf{X}}(\mathbf{x}) & = \prod_{i=1}^{n} f_{X_i}(x_i) \end{cases}$$

Random vector: moments

Mathematical expectation of a random vector is a vector of mathematical expectations of its components:

$$\mathbb{E}\left[\mathbf{X}\right] = (\mathbb{E}X_1, \dots, \mathbb{E}X_n)^\top$$

Second moments of a random vector are described with **covariance matrix** $\mathbb{V}\mathrm{ar}(\mathbf{X}) = \Sigma$, where

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j)$$

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j) = \mathbb{E}\left[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right]$$

In particular, the diagonal elements are variances: $\Sigma_{ii} = \mathbb{V}\mathrm{ar}(X_i)$.

Random vector: LOTUS

Let $g(\cdot) : \mathbb{R}^2 \to \mathbb{R}$, then

$$\mathbb{E}\left[g(X, Y)\right] = \sum_m \sum_n g(m, n) \mathbb{P}(X = m, Y = n)$$

if X, Y are continuous r.v.s; if they are continuous:

$$\mathbb{E}\left[g(X, Y)\right] = \int\int g(x, y) f_{X,Y}(x, y) \mathrm{d}x \mathrm{d}y$$

Covariance

Covariance of two random variables X and Y is defined as

$$\operatorname{cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])\left(Y - \mathbb{E}\left[Y\right]\right)\right] = \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$$

Properties of covariance:

- $\operatorname{cov}(X, X) = \mathbb{V}\text{ar}\left(X\right)$
- $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$
- $\operatorname{cov}(X, c) = 0, \ \forall c$
- $\operatorname{cov}(aX, Y) = a \operatorname{cov}(X, Y)$
- $\operatorname{cov}(X + Y, Z) = \operatorname{cov}(X, Z) + \operatorname{cov}(Y, Z)$
- $\operatorname{cov}(X + Y, Z + W) = \operatorname{cov}(X, Z) + \operatorname{cov}(X, W) + \operatorname{cov}(Y, Z) + \operatorname{cov}(Y, W)$
- $\mathbb{V}\text{ar}\left(X + Y\right) = \mathbb{V}\text{ar}\left(X\right) + \mathbb{V}\text{ar}\left(Y\right) + 2\operatorname{cov}(X, Y)$

Correlation

If $X \perp Y$, $\operatorname{cov}(X, Y) = 0$. The converse is not true. Regardless, covariance is often used to measure the dependency between random variables. It is not handy to use, so instead a **correlation coefficient** is proposed:

$$r_{XY} = \frac{\operatorname{cov}(X, Y)}{\sqrt{\mathbb{V}\text{ar}\left(X\right)\mathbb{V}\text{ar}\left(Y\right)}}$$

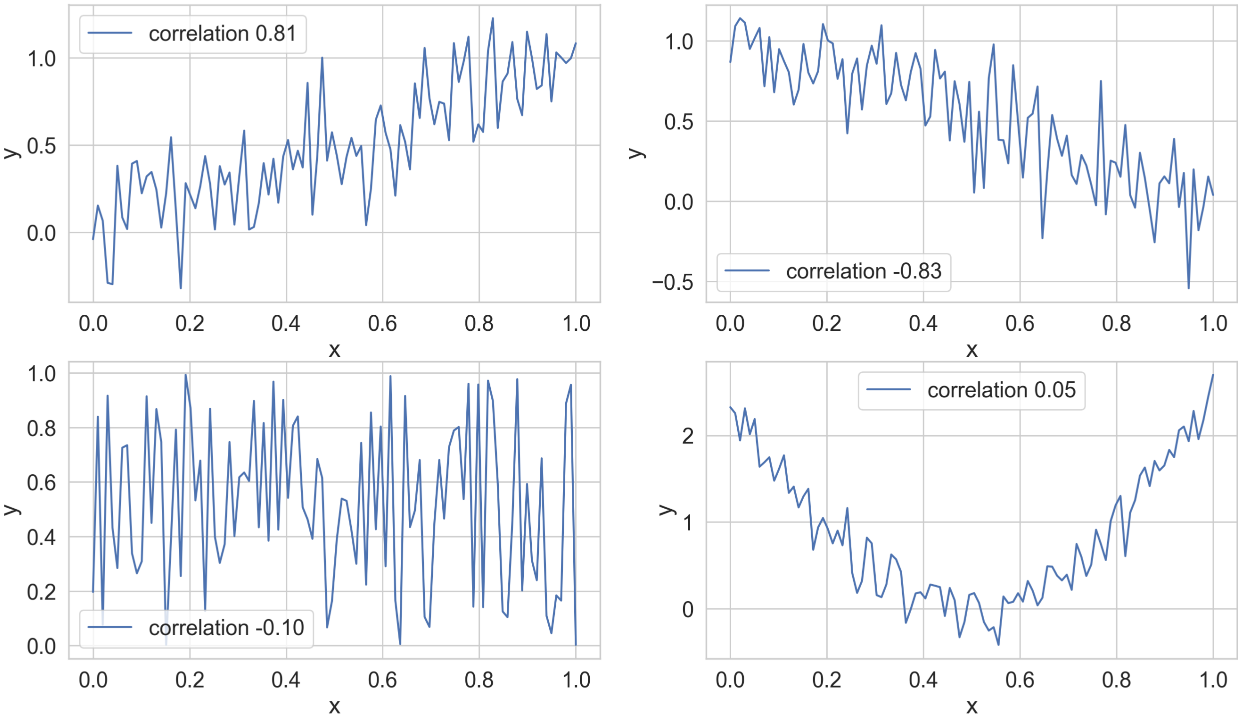
Note that $-1 \leqslant r_{XY} \leqslant 1$.

In [1]:
```
import numpy as np
import scipy.stats as sts

import IPython.display as dp
import matplotlib.pyplot as plt
import seaborn as sns

dp.set_matplotlib_formats("retina")
sns.set(style="whitegrid", font_scale=1.5)
sns.despine()

%matplotlib inline
```

```
/var/folders/33/j0cl7y453td68qb96j7bqcj4cf41kc/T/ipykernel_99751/260074960
0.py:8: DeprecationWarning: `set_matplotlib_formats` is deprecated since I
Python 7.23, directly use `matplotlib_inline.backend_inline.set_matplotlib
_formats()`
  dp.set_matplotlib_formats("retina")
<Figure size 640x480 with 0 Axes>
```

In [4]:
```
n = 100
s = 1/5
x = np.linspace(0, 1, n)
```

```python
y1 = x + sts.norm(scale=s).rvs(n)
y2 = 1 - x + sts.norm(scale=s).rvs(n)
y3 = sts.uniform().rvs(n)
y4 = 10 * (x - 0.5) ** 2 + sts.norm(scale=s).rvs(n)

fig, axes = plt.subplots(2,2,figsize=(16,9))
for ax, y in zip(axes.flatten(), [y1, y2, y3, y4]):
    ax.plot(x, y, label=f"correlation {np.corrcoef(x, y)[0, 1]:.2f}")
    ax.set_xlabel("x")
    ax.set_ylabel("y")
    ax.legend();
```



Random vector: covariance matrix

Matrix notation for covariance matrix is $\mathbb{V}\mathrm{ar}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top\right]$.

Properties of convariance matrix:

- Symmetry $\Sigma^\top = \Sigma$
- Non-negative semi-definite: $a^\top \Sigma a \geqslant 0, \forall a$

Random vector: marginal and conditional distributions

Marginal distribution is the distribution of a subset of a random vector. For example, consider r.v. $\mathbf{X} \in \mathbb{R}^n$ and let's view it as two vectors, $\mathbf{Y} \in \mathbb{R}^k$ and $\mathbf{Z} \in \mathbb{R}^{n-k}$, stacked: $\mathbf{X} = (\mathbf{Y}^\top, \mathbf{Z}^\top)^\top$. The marginal distribution of $\mathbf{Z}$ then will be:

$$f_{\mathbf{Z}}(\mathbf{z}) = \int_{\mathbb{R}^k} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}) d\mathbf{y}$$

In words, we take distribution of $\mathbf{X}$ and **integrate out** everything not realted to $\mathbf{Z}$.

We may also define **conditional distribution**:

$$f_{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(\mathbf{y}) = \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})}$$

Example 4: joint, marginal and conditional distributions for discrete case

Let X be the indicator of the sampled individual being a current smoker, and let Y be the indicator of his developing lung cancer at some point in his life. Suppose the joint PMF is as follows:

	$Y=1$	$Y=0$
$X=1$	$\frac{5}{100}$	$\frac{20}{100}$
$X=0$	$\frac{3}{100}$	$\frac{72}{100}$

Find the marginal and consitional distributions.

Solution 4

	$Y=1$	$Y=0$
$X=1$	$\frac{5}{100}$	$\frac{20}{100}$
$X=0$	$\frac{3}{100}$	$\frac{72}{100}$

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y)$$

	$Y=1$	$Y=0$	Sum
$X=1$	$\frac{5}{100}$	$\frac{20}{100}$	$\frac{25}{100}$
$X=0$	$\frac{3}{100}$	$\frac{72}{100}$	$\frac{75}{100}$
Sum	$\frac{8}{100}$	$\frac{92}{100}$	$\frac{100}{100}$

Solution 4

	$Y=1$	$Y=0$
$X=1$	$\frac{5}{100}$	$\frac{20}{100}$
$X=0$	$\frac{3}{100}$	$\frac{72}{100}$

$$\mathbb{P}(Y=y|X=x)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)}$$

Example: if the person is a smoker ($X=1$), then
$\mathbb{P}(Y=1|X=1)=\frac{\mathbb{P}(X=1,Y=1)}{\mathbb{P}(X=1)}=\frac{5/100}{25/100}=0.2.$

Example 5 (unit disc)

Consider a random point on unit disc with random coordinates (X,Y). What is the joint, marginal and conditional PDF for the coordinates?

Solution 5

The joint is

$$f_{X,Y}(x,y)=\begin{cases}\frac{1}{\pi}, \text{ if } x^2+y^2\leqslant 1,\\ 0, \text{ else}\end{cases}$$

The marginal for X is:

$$f_X(x)=\int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)dy=\int\limits_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{2}{\pi}\sqrt{1-x^2}$$

The conditional for Y is:

$$f_{Y|X=x}(x)=\frac{f_{X,Y}(x,y)}{f_X(x)}=\frac{\frac{1}{\pi}}{\frac{2}{\pi}\sqrt{1-x^2}}=\frac{1}{2\sqrt{1-x^2}}$$