Lecture 5: Moment Generating Functions

IB Paper 7: Probability and Statistics

Carl Edward Rasmussen

Department of Engineering, University of Cambridge

February 28th, 2018

Moment Generating Functions

The computation of the central moments (e.g. expectation and variance) as well as combinations of random variables such as sums are useful, but can be tedious because of the sums or integrals involved.

Example: The expectation of the Binomial is

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{r=0}^{n} r\, p(X=r) = \sum_{r=0}^{n} r\, {}_nC_r\, p^r(1-p)^{n-r} = \sum_{r=1}^{n} \frac{rn!}{(n-r)!r!} p^r(1-p)^{n-r} \\
&= np \sum_{r=1}^{n} \frac{(n-1)!}{(n-r)!(r-1)!} p^{r-1}(1-p)^{n-r} \\
&= np \sum_{\tilde{r}=0}^{\tilde{n}} \frac{\tilde{n}!}{(\tilde{n}-\tilde{r})!\tilde{r}!} p^{\tilde{r}}(1-p)^{\tilde{n}-\tilde{r}} = np,
\end{aligned}$$

where $\tilde{n} = n-1$ and $\tilde{r} = r-1$, and using the fact that the Binomial normalizes to one.

Moment Generating functions are a neat mathematical trick which sometimes sidesteps these tedious calculations.

The Discrete Moment Generating Function

For a discrete random variable, we define the moment generating function

$$g(z) \;=\; \sum_r z^r p(r).$$

This is useful, since when differentiated w.r.t. z an extra factor r appears in the sum, thus

$$g'(z) \;=\; \sum_r r z^{r-1} p(r), \quad \text{and} \quad g''(z) \;=\; \sum_r r(r-1) z^{r-2} p(r).$$

So

$$g'(1) \;=\; \sum_r r p(r), \quad \text{and} \quad g''(1) \;=\; \sum_r (r^2 - r) p(r),$$

and

$$\mathbb{E}[R] \;=\; g'(1), \quad \text{and} \quad \mathbb{E}[R^2] \;=\; g''(1) + g'(1).$$

The Binomial Distribution

The Binomial has

$$g(z) = \sum_r {}_nC_r\, z^r p^r (1-p)^{n-r} = \sum_r {}_nC_r (pz)^r (1-p)^{n-r} = (q+pz)^n,$$

by the Binomial theorem, where we have defined $q = 1 - p$.

Thus, we have

$$g'(z) = np(q+pz)^{n-1}, \quad \text{and} \quad g''(z) = n(n-1)p^2(q+pz)^{n-2}.$$

So

$$g'(1) = np, \quad \text{and} \quad g''(1) = n(n-1)p^2,$$

and

$$\mathbb{E}[X] = np, \quad \text{and} \quad \mathbb{E}[X^2] = n^2p^2 - np^2 + np,$$

which combine to

$$\mathbb{E}[X] = np, \quad \text{and} \quad \mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = np - np^2 = npq.$$

Some Discrete Moment Generating Functions

distribution	symbol	probability	moment generating function
Bernoulli	$\text{Ber}(p)$	$p(x) = p^x(1-p)^{1-x}$	$g(z) = q + zp$
Binomial	$B(n, p)$	$p(r) = {}_nC_r\, p^r(1-p)^{n-r}$	$g(z) = (q + zp)^n$
Poisson	$\text{Po}(\lambda)$	$p(r) = \exp(-\lambda)\lambda^r/r!$	$g(z) = \exp(\lambda(z-1))$

where we have defined $q = 1 - p$.

Sums of Random Variables

Example: Consider $W = X + Y$, where $X \sim \text{Po}(\lambda_x)$ and $Y \sim \text{Po}(\lambda_y)$ are independent Poisson distributed. Then

$$
\begin{aligned}
&p(W = w) \\
&= \sum_{x \leqslant w} p(X = x)p(Y = w - x) \; = \; \sum_{x=0}^{w} \exp(-\lambda_x)\frac{\lambda_x^x}{x!}\exp(-\lambda_y)\frac{\lambda_y^{w-x}}{(w-x)!} \\
&= \frac{\exp(-\lambda_x - \lambda_y)}{w!}\sum_{x=0}^{w}\frac{w!}{x!(w-x)!}\lambda_x^x\lambda_y^{w-x} \; = \; \exp(-\lambda_x - \lambda_y)\frac{(\lambda_x + \lambda_y)^w}{w!} \\
&= \text{Po}(\lambda_x + \lambda_y),
\end{aligned}
$$

i.e. the Poisson distribution is closed under addition.

Sums using Moment Generating Functions

Now $W = X + Y$, then

$$\begin{aligned} g_w(z) &= \sum_w z^w \sum_x p(X=x)p(Y=w-x) \\ &= \sum_w \sum_x z^x p(X=x) z^{w-x} p(Y=w-x) \\ &= \sum_x \sum_y z^x p(X=x) z^y p(Y=y) \\ &= \sum_x z^x p(X=x) \sum_y z^y p(Y=y) \\ &= g_x(z) g_y(z). \end{aligned}$$

I.e., the sum of independent random variables has a moment generating function, which is the product of the moment generating functions.

Example: we see immediately, that the sum of two independent Poisson is Poisson with $\lambda = \lambda_x + \lambda_y$ as $g(z) = \exp(\lambda(z-1))$.

Moment Generating Functions in the Continuous case

For continuous distributions[1]

$$g(s) \;=\; \int_x \exp(sx)p(x)dx,$$

which is related to the two-sided Laplace transform. We have

$$g'(s) \;=\; \int x\exp(sx)p(x)dx, \quad \text{and} \quad g''(s) \;=\; \int x^2\exp(sx)p(x)dx,$$

and so on, which gives

$$\mathbb{E}[X] \;=\; g'(0), \quad \text{and} \quad E[X^2] \;=\; g''(0).$$

Also, the sum of two independent continuous random variables, which is the convolution of the probability densities, has a moment generating function which is the product of the moment generating functions.

Similar to the discrete case and to Laplace transforms from signal analysis.

[1]In the past a different definition $g(s) = \int_x \exp(-sx)p(x)dx$ was used.

Moment Generating Functions in the Continuous case

distribution	symbol	probability	moment generating function
Uniform	$\text{Uni}(a, b)$	$p(x) = 1/(b-a)$	$g(s) = \frac{\exp(as)-\exp(bs)}{s(b-a)}$
Exponential	$\text{Ex}(\lambda)$	$p(x) = \lambda\exp(-\lambda x)$	$g(s) = \lambda/(\lambda - s)$
Gaussian	$N(\mu, \sigma^2)$	$p(x) = \frac{\exp(-\frac{(x-\mu)^2}{2\sigma^2})}{\sqrt{2\pi\sigma^2}}$	$g(s) = \exp(s\mu + s^2\sigma^2/2)$

The moment generating functions for shifted and scaled random variables are

$$Y = X + \beta, \qquad g_y(s) = \exp(\beta s) g_x(s)$$

and

$$Y = \alpha X, \qquad g_y(s) = g_x(\alpha s),$$

which are both verified by plugging into the definition.

The multivariate Gaussian

The multivariate Gaussian in D dimensions, where $\mathbf{x}$ is a vector of length D has probability density

$$p(\mathbf{x}) \;\sim\; \mathcal{N}(\boldsymbol{\mu}, \Sigma) \;=\; (2\pi)^{-D/2}|\Sigma|^{-1/2}\exp\big(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\big),$$

where $\boldsymbol{\mu}$ is the mean vector of length D and Σ is the $D \times D$ covariance matrix.

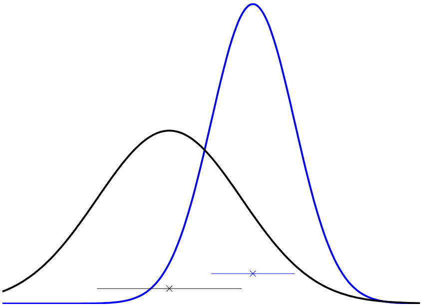
The covariance matrix is positive definite and symmetric.

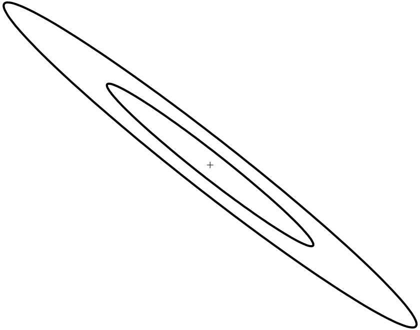
The entries of the covariance matrix Σ_{ij} are the covariances between different coordinates of $\mathbf{x}$

$$\Sigma_{ij} \;=\; \mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)] \;=\; \mathbb{E}[x_i x_j] - \mu_i \mu_j.$$

In a Gaussian, if all covariances $\Sigma_{i \neq j}$ are zero, Σ is diagonal, and the components x_i are independent, since then $p(\mathbf{x}) = \prod_i p(x_i)$.

The Gaussian Distribution





In the multivariate Gaussian, the equi-probability contours are ellipses. The axis directions are given by the eigenvectors of the covariance matrix and their lengths are proportional to the square root of the corresponding eigenvalues.

Correlation and independence

The covariance matrix is sometimes written as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

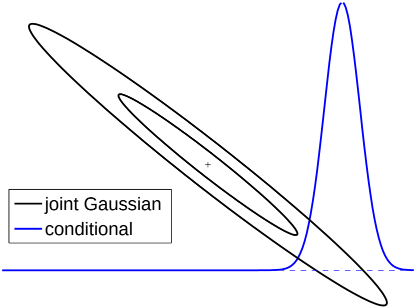
where $-1 < \rho < 1$ is the correlation coefficient. When

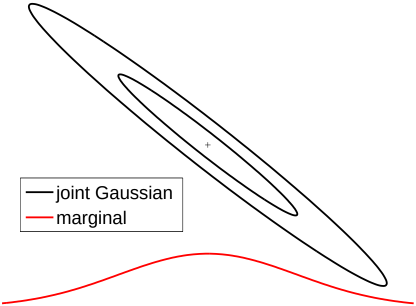
- $\rho < 0$, the variables are anti-correlated
- $\rho = 0$, uncorrelated
- $\rho > 0$, positively correlated

Independence: $p(X, Y) = p(X)p(Y)$. Note: independence $\Rightarrow$ uncorrelated, *but not vice versa*.

Example: X_i are independent, with $X_i \sim N(0, 1)$ and $Y_i = \pm X_i$ (with random sign). Here, X and Y are uncorrelated, *but not independent*.

Conditionals and Marginals of a Gaussian





Both the conditionals and the marginals of a joint Gaussian are again Gaussian.

Conditionals and Marginals of a Gaussian

Recall marginalization:

$$p(\mathbf{x}) \;=\; \int p(\mathbf{x},\mathbf{y})d\mathbf{y}.$$

For Gaussians:

$$p(\mathbf{x},\mathbf{y}) \;=\; N\Big(\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}, \begin{bmatrix}A & C\\ C^\top & B\end{bmatrix}\Big) \;\Longrightarrow\; p(\mathbf{x}) \;=\; \mathcal{N}(\mathbf{a},\ A).$$

And conditioning

$$p(\mathbf{x}|\mathbf{y}) \;=\; N\big(\mathbf{a}+CB^{-1}(\mathbf{y}-\mathbf{b}),\ A-CB^{-1}C^\top\big).$$

The Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ are all identically independently distributed random variables with mean μ and variance σ^2, then in the limit of large n

$$X_1 + X_2 + \ldots + X_n \sim N(n\mu, n\sigma^2),$$

regardless of the actual distribution of X_i. Note: As we expect, the means and the variances add up.

Equivalently

$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \sim N(0, 1).$$

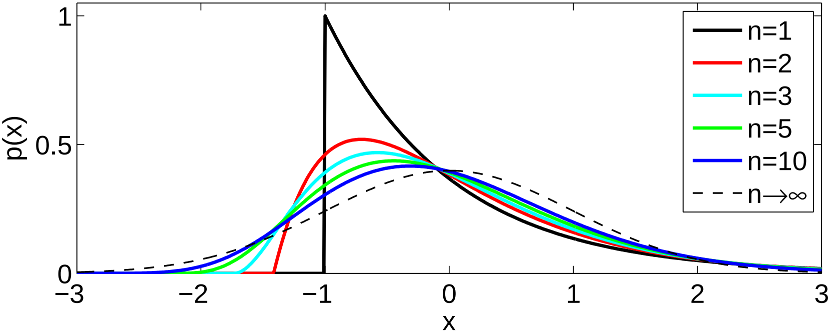
The Central Limit Theorem can be proven by examining the moment generating function.

Central Limit Theorem Example

The distribution of

$$X_n = \frac{Y_1 + Y_2 + \ldots + Y_n - n\mu}{\sigma\sqrt{n}}$$

where $Y_i \sim \text{Ex}(1)$ for different values of n



Even for quite small values of n we get a good approximation by the Gaussian.