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

\[
\mathbb{E}[X] = \sum_{r=0}^{n} r \ p(X = r) = \sum_{r=0}^{n} r \ nC_r \ p^r (1 - p)^{n-r} = \sum_{r=1}^{n} \frac{rn!}{(n-r)!r!} p^r (1 - p)^{n-r}
\]

\[
= np \sum_{r=1}^{n} \frac{(n-1)!}{(n-r)!(r-1)!} p^{r-1} (1 - p)^{n-r}
\]

\[
= np \sum_{\tilde{r}=0}^{\tilde{n}} \frac{\tilde{n}!}{(\tilde{n}-\tilde{r})!\tilde{r}!} p^{\tilde{r}} (1 - p)^{\tilde{n}-\tilde{r}} = np,
\]

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

\[ g'(z) = \sum_{r} rz^{r-1} p(r), \quad \text{and} \quad g''(z) = \sum_{r} r(r-1)z^{r-2} p(r). \]

So

\[ g'(1) = \sum_{r} rp(r), \quad \text{and} \quad g''(1) = \sum_{r} (r^2 - r)p(r), \]

and

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

The Binomial has

\[ g(z) = \sum_{r} \binom{n}{r} z^r p^r (1 - p)^{n-r} = \sum_{r} \binom{n}{r} (pz)^r (1 - p)^{n-r} = (q + pz)^n, \]

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

\[ \mathbb{E}[X] = np, \quad \text{and} \quad \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = np - np^2 = npq. \]
### Some Discrete Moment Generating Functions

<table>
<thead>
<tr>
<th>distribution</th>
<th>symbol</th>
<th>probability</th>
<th>moment generating function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli</td>
<td>Ber(p)</td>
<td>( p(x) = p^x(1-p)^{1-x} )</td>
<td>( g(z) = q + zp )</td>
</tr>
<tr>
<td>Binomial</td>
<td>B(n, p)</td>
<td>( p(r) = \binom{n}{r} p^r (1-p)^{n-r} )</td>
<td>( g(z) = (q + zp)^n )</td>
</tr>
<tr>
<td>Poisson</td>
<td>Po(( \lambda ))</td>
<td>( p(r) = \exp(-\lambda)\lambda^r / r! )</td>
<td>( g(z) = \exp(\lambda(z-1)) )</td>
</tr>
</tbody>
</table>

where we have defined \( q = 1 - p \).
**Sums of Random Variables**

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

\[
p(Z = z) = \sum_{x \leq z} P(X = x)P(Y = z - x) = \sum_{x=0}^{z} \exp(-\lambda_x) \frac{\lambda_x^x}{x!} \exp(-\lambda_y) \frac{\lambda_y^{z-x}}{(z-x)!}
\]

\[
= \frac{\exp(-\lambda_x - \lambda_y)}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda_x^x \lambda_y^{z-x} = \exp(-\lambda_x - \lambda_y) \frac{(\lambda_x + \lambda_y)^z}{z!}
\]

\[
= \text{Po}(\lambda_x + \lambda_y),
\]

i.e. the Poisson distribution is closed under addition.
Sums using Moment Generating Functions

Now $W = X + Y$, then

$$g_W(z) = \sum_{w} z^w \sum_{x} p(X = x)p(Y = w - x)$$

$$= \sum_{w} \sum_{x} z^x p(X = x)z^{w-x} p(Y = w - x)$$

$$= \sum_{x} \sum_{y} z^x p(X = x)z^y p(Y = y)$$

$$= \sum_{x} z^x p(X = x) \sum_{y} z^y p(Y = y)$$

$$= g_x(z)g_y(z).$$

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

\[ g(s) = \int_x \exp(sx)p(x)dx, \]

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

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

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.

---

1In the past a different definition \( g(s) = \int_x \exp(-sx)p(x)dx \) was used.
Moment Generating Functions in the Continuous case

<table>
<thead>
<tr>
<th>distribution</th>
<th>symbol</th>
<th>probability</th>
<th>moment generating function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>Uni(a, b)</td>
<td>( p(x) = 1/(b - a) )</td>
<td>( g(s) = \frac{\exp(as) - \exp(bs)}{s(b-a)} )</td>
</tr>
<tr>
<td>Exponential</td>
<td>Ex(( \lambda ))</td>
<td>( p(x) = \lambda \exp(-\lambda x) )</td>
<td>( g(s) = \frac{\lambda}{(\lambda - s)} )</td>
</tr>
<tr>
<td>Gaussian</td>
<td>N(( \mu ), ( \sigma^2 ))</td>
<td>( p(x) = \frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} )</td>
<td>( g(s) = \exp(s\mu + \frac{s^2\sigma^2}{2}) )</td>
</tr>
</tbody>
</table>

The moment generating functions for shifted and scaled random variables are

\[
Y = X + \beta, \quad g_Y(s) = \exp(\beta s) g_X(s)
\]

and

\[
Y = \alpha X, \quad g_Y(s) = g_X(\alpha s),
\]

which are both verified by plugging into the definition.
The multivariate Gaussian in $D$ dimensions, where $x$ is a vector of length $D$ has probability density

$$p(x) \sim N(\mu, \Sigma) = (2\pi)^{-D/2}|\Sigma|^{-1/2} \exp \left( -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right),$$

where $\mu$ is the mean vector of length $D$ and $\Sigma$ is the $D \times D$ covariance matrix. The covariance matrix is positive definite and symmetric.

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

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

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

Rasmussen (CUED) 
Lecture 5: Moment Generating Functions 
March 1st, 2017
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

\[
Σ = \begin{bmatrix}
σ_1^2 & ρσ_1σ_2 \\
ρσ_1σ_2 & σ_2^2
\end{bmatrix},
\]

where \(-1 < ρ < 1\) is the correlation coefficient. When

- \(ρ < 0\), the variables are anti-correlated
- \(ρ = 0\), uncorrelated
- \(ρ > 0\), positively correlated

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

Example: \(X_i\) are independent, with \(X_i \sim N(0, 1)\) and \(Y_i = ±X_i\) (with random sign). Here, \(X\) and \(Y\) are uncorrelated, but not independent.
Both the \textbf{conditionals} and the \textbf{marginals} of a joint Gaussian are again Gaussian.
Recall marginalization:
\[ p(x) = \int p(x, y) \, dy. \]

For Gaussians:
\[ p(x, y) = N \left( \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \right) \implies p(x) = N(a, A). \]

And conditioning
\[ p(x|y) = N \left( a + CB^{-1}(y - b), A - CB^{-1}C^\top \right). \]
The Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ are all identically independently distributed random variables with mean $\mu$ and variance $\sigma^2$, the 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.