Multinomial Distribution

- The Binomial distribution can be extended to describe number of outcomes in a series of independent trials each having more than 2 possible outcomes.

- If a given trial can result in the \( k \) outcomes \( E_1, E_2, \ldots, E_k \) with probabilities \( p_1, p_2, \ldots, p_k \), then the probability distribution of the random variables \( X_1, X_2, \ldots, X_k \), representing the number of occurrences for \( E_1, E_2, \ldots, E_k \) in \( n \) independent trials is

\[
p_{x_1, \ldots, x_k} (x_1, \ldots, x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}
\]

with \( \sum_{i=1}^{k} x_i = n \), and \( \sum_{i=1}^{k} p_i = 1 \).
Properties of Expectations Involving Joint Distributions

• For random variables $X, Y$ and constants $a, b \in R$

$$E(aX + bY) = aE(X) + bE(Y)$$

Proof:

• For independent random variables $X, Y$

$$E(XY) = E(X)E(Y)$$

whenever these expectations exist.

Proof:
Covariance

• Recall: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{E}[(X-\text{E}(X))(Y-\text{E}(Y))]$

• Definition
  For random variables $X, Y$ with $E(X), E(Y) < \infty$, the covariance of $X$ and $Y$ is
  
  $\text{Cov}(X, Y) = \text{E}[(X - \text{E}(X))(Y - \text{E}(Y))]$

• Covariance measures whether or not $X-\text{E}(X)$ and $Y-\text{E}(Y)$ have the same sign.

• Claim:
  $\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y)$

  Proof:

• Note: If $X, Y$ independent then $E(XY) = E(X)E(Y)$, and $\text{Cov}(X, Y) = 0$. 
Example

• Suppose $X, Y$ are discrete random variables with probability function given by

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>1/8</td>
<td>1/8</td>
</tr>
<tr>
<td>0</td>
<td>1/8</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1/8</td>
<td>1/8</td>
</tr>
</tbody>
</table>

• Find Cov$(X,Y)$. Are $X,Y$ independent?
Important Facts

• Independence of $X, Y$ implies $\text{Cov}(X,Y) = 0$ but NOT vice versa.

• If $X, Y$ independent then $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

• If $X, Y$ are NOT independent then

\[ \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y). \]

• $\text{Cov}(X,X) = \text{Var}(X)$. 
Properties of Covariance

For random variables $X, Y, Z$ and constants $a, b, c, d \in R$

- $\text{Cov}(aX+b, cY+d) = ac\text{Cov}(X,Y)$
- $\text{Cov}(X+Y, Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$
- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$
Correlation

• **Definition**
For $X$, $Y$ random variables the *correlation* of $X$ and $Y$ is

$$
\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{V(X)} \sqrt{V(Y)}}
$$

whenever $V(X)$, $V(Y) \neq 0$ and all these quantities exists.

• **Claim:**

$$
\rho(aX+b, cY+d) = \rho(X,Y)
$$

**Proof:**

• This claim means that the correlation is scale invariant.
Theorem

- For $X, Y$ random variables, whenever the correlation $\rho(X,Y)$ exists it must satisfy

\[-1 \leq \rho(X,Y) \leq 1\]

Proof:
Interpretation of Correlation $\rho$

- $\rho(X,Y)$ is a measure of the strength and direction of the linear relationship between $X$, $Y$.

- If $X$, $Y$ have non-zero variance, then $\rho \in [-1, 1]$. 

- If $X$, $Y$ independent, then $\rho(X,Y) = 0$. Note, it is not the only time when $\rho(X,Y) = 0$ !!!

- $Y$ is a linearly increasing function of $X$ if and only if $\rho(X,Y) = 1$.

- $Y$ is a linearly decreasing function of $X$ if and only if $\rho(X,Y) = -1$. 
Example

• Find \( \text{Var}(X - Y) \) and \( \rho(X,Y) \) if \( X, Y \) have the following joint density

\[
f_{x,y}(x, y) = \begin{cases} 
3x & 0 \leq y \leq x \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]
Generating Functions

• For a sequence of real numbers \( \{a_j\} = a_0, a_1, a_2, \ldots, \) the generating function of \( \{a_j\} \) is

\[
A(t) = \sum_{j=0}^{\infty} a_j t^j
\]

if this converges for \(|t| < t_0\) for some \(t_0 > 0\).
Probability Generating Functions

• Suppose $X$ is a random variable taking the values 0, 1, 2, … (or a subset of the non-negative integers).

• Let $p_j = P(X = j)$, $j = 0, 1, 2, …$. This is in fact a sequence $p_0, p_1, p_2, …$

• **Definition:** The *probability generating function* of $X$ is

$$
\pi_X(t) = p_0 + p_1 t + p_2 t^2 + \cdots = \sum_{j=0}^{\infty} p_j t^j
$$

• Since $|p_j t^j| \leq p_j$ if $|t| < 1$ and $\sum_{j=0}^{\infty} p_j = 1$ the pgf converges absolutely at least for $|t| < 1$.

• In general, $\pi_X(1) = p_0 + p_1 + p_2 + \cdots = 1$.

• The pgf of $X$ is expressible as an expectation:

$$
\pi_X(t) = \sum_{j=0}^{\infty} p_j t^j = E(t^X)
$$
Examples

- $X \sim \text{Binomial}(n, p)$,

$$
\pi_X(t) = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} t^j = (pt + q)^n
$$

converges for all real $t$.

- $X \sim \text{Geometric}(p)$,

$$
\pi_X(t) = \sum_{j=1}^{\infty} pq^{j-1} t^j = \frac{pt}{1 - qt}
$$

converges for $|qt| < 1$ i.e. $|t| < \frac{1}{\frac{1}{q}} = \frac{1}{1 - p}$

Note: in this case $p_j = pq^j$ for $j = 1, 2, \ldots$
PGF for sums of independent random variables

• If $X, Y$ are independent and $Z = X + Y$ then,

$$
\pi_Z(t) = E(t^Z) = E(t^{X+Y}) = E(t^X t^Y) = E(t^X)E(t^Y) = \pi_X(t)\pi_Y(t)
$$

• Example

Let $Y \sim \text{Binomial}(n, p)$. Then we can write $Y = X_1 + X_2 + \ldots + X_n$. Where $X_i$’s are i.i.d Bernoulli($p$). The pgf of $X_i$ is

$$
\pi_{X_i}(t) = t^0(1 - p) + t^1 p = tp + q.
$$

The pgf of $Y$ is then

$$
\pi_Y(t) = E(t^{X_1+X_2+\ldots+X_n}) = E(t^{X_1})E(t^{X_2})\ldots E(t^{X_n}) = (tp + q)^n.
$$
Use of PGF to find probabilities

• **Theorem**

Let $X$ be a discrete random variable, whose possible values are the nonnegative integers. Assume $\pi_X(t_0) < \infty$ for some $t_0 > 0$. Then

$$\pi_X(0) = P(X = 0),$$

$$\pi'_X(0) = P(X = 1),$$

$$\pi''_X(0) = 2P(X = 2),$$

etc. In general,

$$\pi^{(k)}_X(0) = k! P(X = k),$$

where $\pi^{(k)}_X$ is the $k^{th}$ derivative of $\pi_X$ with respect to $t$.

• **Proof:**
Example

- Suppose $X \sim \text{Poisson}(\lambda)$. The pgf of $X$ is given by

$$\pi_X(t) = \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} t^j =$$

- Using this pgf we have that
Finding Moments from PGFs

• **Theorem**
  Let $X$ be a discrete random variable, whose possible values are the nonnegative integers. If $\pi_X(t) < \infty$ for $|t| < t_0$ for some $t_0 > 1$. Then
  
  $\pi'_X(1) = E(X)$,
  
  $\pi''_X(1) = E(X(X-1))$,
  
  etc. In general,
  
  $\pi^{(k)}_X(1) = E(X(X-1)(X-2)\cdots(X-K+1))$,

  Where $\pi^{(k)}_X$ is the $k$th derivative of $\pi_X$ with respect to $t$.

• **Note**: $E(X(X-1)\cdots(X-k+1))$ is called the $k$th factorial moment of $X$.

• **Proof:**
Example

• Suppose $X \sim \text{Binomial}(n, p)$. The pgf of $X$ is

$$\pi_X(t) = (pt+q)^n.$$ 

Find the mean and the variance of $X$ using its pgf.
Uniqueness Theorem for PGF

- Suppose $X$, $Y$ have probability generating function $\pi_X$ and $\pi_Y$ respectively. Then $\pi_X(t) = \pi_Y(t)$ if and only if $P(X = k) = P(Y = k)$ for $k = 0, 1, 2, \ldots$

- **Proof:**
  Follow immediately from calculus theorem:
  If a function is expressible as a power series at $x=a$, then there is only one such series.
  A pgf is a power series about the origin which we know exists with radius of convergence of at least 1.
Moment Generating Functions

• The *moment generating function* of a random variable $X$ is

$$m_X(t) = E(e^{tx})$$

$m_X(t)$ exists if $m_X(t) < \infty$ for $|t| < t_0 > 0$

• If $X$ is discrete

$$m_X(t) = \sum_x e^{tx} p_X(x).$$

• If $X$ is continuous

$$m_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx.$$

• Note: $m_X(t) = \pi_X(e^t)$. 
Examples

• \( X \sim \text{Exponential}(\lambda) \). The mgf of \( X \) is

\[
m_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} \, dx =
\]

• \( X \sim \text{Uniform}(0,1) \). The mgf of \( X \) is

\[
m_X(t) = E(e^{tX}) = \int_0^1 e^{tx} \, dx =
\]
Generating Moments from MGFs

- **Theorem**
  Let $X$ be any random variable. If $m_X(t) < \infty$ for $|t| < t_0$ for some $t_0 > 0$. Then
  
  \begin{align*}
  m_X(0) &= 1 \\
  m'_X(0) &= E(X), \\
  m''_X(0) &= E(X^2),
  \end{align*}
  
  etc. In general,
  
  \begin{align*}
  m_X^{(k)}(0) &= E(X^k),
  \end{align*}
  
  Where $m_X^{(k)}$ is the $k$th derivative of $m_X$ with respect to $t$.

- **Proof:**
Example

- Suppose $X \sim \text{Exponential}(\lambda)$. Find the mean and variance of $X$ using its moment generating function.
Example

• Suppose $X \sim N(0,1)$. Find the mean and variance of $X$ using its moment generating function.
Properties of Moment Generating Functions

• $m_X(0) = 1$.

• If $Y=a+bX$, $a, b \in \mathbb{R}$ then the mgf of $Y$ is given by

$$m_Y(t) = E(e^{tY}) = E(e^{at+btX}) = e^{at} E(e^{btX}) = e^{at} m_X(bt).$$

• If $X, Y$ independent and $Z = X+Y$ then,

$$m_Z(t) = E(e^{tZ}) = E(e^{tX+tY}) = E(e^{tX} e^{tY}) = E(e^{tX}) E(e^{tY}) = m_X(t)m_Y(t).$$
Uniqueness Theorem

• If a moment generating function $m_X(t)$ exists for $t$ in an open interval containing 0, it uniquely determines the probability distribution.
Example

• Find the mgf of \( X \sim N(\mu, \sigma^2) \) using the mgf of the standard normal random variable.

• Suppose, \( X_1 \sim N(\mu_1, \sigma_1^2) \), \( X_2 \sim N(\mu_2, \sigma_2^2) \) independent.

Find the distribution of \( X_1 + X_2 \) using mgf approach.
Characteristic Function

• The characteristic function, $c_X$, of a random variable $X$ is defined by:

$$c_X(t) = E(e^{isX})$$

for $s \in R$.

• The definition of the characteristics function is just like the definition of the mgf, except for the introduction of the imaginary number $i = \sqrt{-1}$.

• Using properties of complex numbers, we see that the characteristic function can also be written as

$$c_X(s) = E(\cos(sX)) + iE(\sin(sX))$$

for $s \in R$.

• The characteristics function unlike the mgf is always finite…
Properties of Characteristic Function

- The characteristics function has many nice properties similar to the mgf.
- In particular, it can be used to generate moments.

- Let $X$ be any random variable with its first $k$ moments finite. Then

  $$c_X(0) = 1$$
  $$c_X'(0) = iE(X),$$
  $$c_X''(0) = i^2 E(X^2) = -E(X^2)$$

  etc. In general,

  $$c_X^{(k)}(0) = i^k E(X^k),$$

  Where $c_X^{(k)}$ is the $k$th derivative of $c_X$ with respect to $s$.

- Let $X$ and $Y$ be independent random variables. Then

  $$c_{X+Y}(s) = c_X(s) c_Y(s)$$
Examples