STAT:5100 (22S:193) Statistical Inference I

Week 8

Luke Tierney

University of Iowa

Fall 2015

Recap

- Background on integration
- Binomial distribution
- Negative binomial distribution
- Hypergeometric distribution

Poisson Distribution

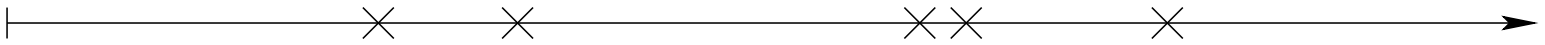
- X has a Poisson distribution if it is discrete with $\mathcal{X} = \{0, 1, 2, \dots\}$ and PMF
$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$
for $x = 0, 1, \dots$ and some $\lambda \geq 0$.
- The MGF and cumulant generating function are
$$M(t) = \exp\{\lambda(e^t - 1)\}$$
$$K(t) = \lambda(e^t - 1)$$
- The mean and variance are
$$E[X] = \frac{d}{dt} \lambda(e^t - 1)\Big|_{t=0} = \lambda$$
$$\text{Var}(X) = \frac{d}{dt}^2 \lambda(e^t - 1)\Big|_{t=0} = \lambda$$

Some plots can be obtained with

```
x <- 0 : 10
plot(x, dpois(x, 0.5), type = "h", lwd = 3)
plot(x, dpois(x, 1), type = "h", lwd = 3)
plot(x, dpois(x, 3), type = "h", lwd = 3)
x <- 0 : 50
plot(x, dpois(x, 10), type = "h", lwd = 3)
plot(x, dpois(x, 30), type = "h", lwd = 3)
plot(x, dpois(x, 20), type = "h", lwd = 3)
plot(x, dpois(x, 30), type = "h", lwd = 3)
```

Poisson Process

The Poisson distribution often occurs in the context of a *Poisson process*: Suppose certain "events" occur at random points in time.



Examples

- Cars arriving at an intersection
- Requests arriving at a network server
- Atoms decaying

Suppose

(i) Only finitely many events occur in a finite interval.

(ii) Events occur one at a time.

(iii) For any s, t the distribution of the number of events in $(s, s+t]$ does not depend on s.

(iv) For any two disjoint time intervals I_1, I_2 the numbers of events in I_1 and I_2 are independent.

Then the process of events is called a *Poisson process*.

- The number of events in an interval of length t has a Poisson distribution with mean λt for some $\lambda > 0$.
- λ is the expected number of events in a period of unit length.
- λ is also called the *event rate* of the process.
- Poisson processes can be extended to model random point patterns in higher dimensions.

Example

- Suppose calls arrive at a switch at a rate of 75 per hour.
- What is the chance that more than 2 calls arrive in a particular one-minute interval?
- A Poisson process may be a reasonable model for the call process.
- We have

$$\begin{aligned} \lambda &= 75 \quad \text{calls/hour} \\ t &= 1/60 \quad \text{hours} \\ \lambda t &= 5/4 \quad \text{calls} \end{aligned}$$

- Therefore the number of calls in the one-minute interval has a Poisson distribution with mean 5/4.

Example (continued)

- The probability of more than two calls is therefore

$$
\begin{aligned}
P(\text{more than two calls}) &= 1 - f(0) - f(1) - f(2) \\
&= 1 - e^{-5/4} - \frac{5}{4}e^{-5/4} - \frac{1}{2}\left(\frac{5}{4}\right)^2 e^{-5/4} \\
&= 1 - \frac{97}{32}e^{-5/4} \\
&\approx 0.1315
\end{aligned}
$$

- Let T_1 be the time of the first event in a Poisson process with rate λ.
- Then for $t > 0$

$$P(T_1 > t) = P(\text{no events in } [0, t]) = e^{-\lambda t}.$$

- So T_1 has an exponential distribution with mean $1/\lambda$.
- If T_k is the time between the $k - 1$-st and k-th event then T_k also has an exponential distribution with mean $1/\lambda$.
- The times $T_1, T_2, \ldots$ are mutually independent.

Uniform Distribution

- X is uniform on $[a, b]$ if

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

- We have already seen

$$E[X] = \frac{a+b}{2}$$
$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

- Not often (but sometimes) useful as a model.
- Often useful as a test case.

- If $X \sim \text{Uniform}[a, b]$, then

$$Y = \frac{X - a}{b - a} \sim \text{Uniform}[0, 1]$$

- The family of uniform distributions is an example of a *location-scale* family.
- Given a random variable Y with density f_Y one can define a family of distributions for constants a and b as

$$X = aY + b$$

  with densities

$$f_X(x|a, b) = \frac{1}{a} f_Y\left(\frac{x - b}{a}\right)$$

- For many standard parametric families one of the parameters is a location or a scale parameter.

Gamma Distribution

- X has a gamma distribution with parameters $\alpha, \beta > 0$ if

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

- Here

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

- The parameter β is a scale parameter.
- Some plots can be obtained with

```
plot(function(x) dgamma(x, 0.5), 0, 5)
plot(function(x) dgamma(x, 1), 0, 5)
plot(function(x) dgamma(x, 2), 0, 10)
plot(function(x) dgamma(x, 5), 0, 20)
plot(function(x) dgamma(x, 100), 0, 150)
```

- Some useful properties of $\Gamma(\alpha)$:

$$\begin{aligned} \Gamma(\alpha+1) &= \alpha\Gamma(\alpha) && \text{for all } \alpha > 0 \\ \Gamma(1) &= 1 \\ \Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi} \end{aligned}$$

- It follows that for integers $n \geq 1$

$$\Gamma(n) = (n-1)!$$

- As a result,

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}$$

- This can be useful computationally and can also be used to generalize the negative binomial distribution to non-integer r.

- We have already seen:

$$E[X] = \alpha\beta$$
$$\mathrm{Var}(X) = \alpha\beta^2$$
$$M(t) = \begin{cases} \left(\frac{1}{1-\beta t}\right)^\alpha & t < 1/\beta \\ \infty & \text{otherwise} \end{cases}$$

- For general α the CDF cannot be computed in closed form.
- But the CDF can be computed in terms of the *incomplete gamma function*

$$\Gamma(\alpha, t) = \int_0^t x^{\alpha-1} e^{-x} dx.$$

- Good algorithms for evaluating this are available.

- Special cases:

$\alpha = 1, \beta = 1/\lambda$	Exponential distribution
$\alpha = p/2, \beta = 2$	χ^2_p distribution
$\alpha = n, \beta = 1/\lambda$	Erlang-n distribution

- For the Poisson process, the time of the n-th event has an Erlang-n distribution.

- Let X be Erlang-n.
- Then

$$\begin{aligned} P(X > t) &= P(\text{fewer than } n \text{ events by time } t) \\ &= P(N \leq n-1) \end{aligned}$$

with $N \sim \text{Poisson}(\lambda t)$.

- Therefore

$$P(X > t) = \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

- This can also be shown using integration by parts.
- This is actually more useful in reverse:

$$P(N \leq n-1) = P(X > t)$$

- Since $P(X > t)$ is easy to evaluate numerically, this allows a Poisson CDF to be computed without summing a long series.

Normal Distibution

- X has a normal distribution $N(\mu, \sigma^2)$ if

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

  for all $x \in \mathbb{R}$ where $\sigma > 0$.
- If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

  has a *standard normal distribution*.
- All normal probabilities can be found using standard normal ones.
- Again, closed form expressions do not exist, but good numerical approximations are available.

- Even showing that the standard normal density integrates to one is not easy.
- If we know that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, we can use the change of variables $Y = X^2$ to show this.
- But $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ is usually derived from the normal distribution.
- A direct proof requires a "trick:"

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-z^2/2} dz\right)^2 &= \int_{-\infty}^{\infty} e^{-t^2/2} dt \int_{-\infty}^{\infty} e^{-u^2/2} du \\ &= \int\int e^{-(t^2+u^2)/2} dt\, du\end{aligned}$$

- Change variables to polar coordinates:

$$t = r\cos\theta$$
$$u = r\sin\theta$$

for $0 \le r, 0 \le \theta < 2\pi$.

- The Jacobian is $|r|$, so

$$\begin{aligned}\int\int e^{-(t^2+u^2)/2} dt\, du &= \int_0^\infty \int_0^{2\pi} e^{-r^2/2} r d\theta\, dr \\ &= 2\pi \int_0^\infty r e^{-r^2/2} dr \\ &= 2\pi.\end{aligned}$$

- Therefore

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}.$$

Recap

- Poisson distribution
- Poisson process
- Uniform distribution
- Gamma distribution
- Normal distribution

- The MGF of a standard normal Z is

$$
\begin{aligned}
M_Z(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz - z^2/2} dz \\
&= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} dz \\
&= e^{t^2/2}.
\end{aligned}
$$

- This is finite for all t so all moments exist.
- By symmetry, the mean of a standard normal Z is

$$E[Z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0.$$

- Similarly, $E[Z^k] = 0$ for all positive odd integers k.

- The variance of a standard normal Z is

$$\begin{aligned}
\text{Var}(Z) = E[Z^2] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -z\left(-ze^{-z^2/2}\right) dz\\
&= -\frac{z}{\sqrt{2\pi}}e^{-z^2/2}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-z^2/2} dz = 1
\end{aligned}$$

- Even moments of the standard normal distribution are

$$\begin{aligned}
E[Z^n] &= \int_{-\infty}^{\infty} x^n \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^n e^{-x^2/2} dx\\
&= \frac{2}{\sqrt{2\pi}}\int_0^{\infty} (\sqrt{2y})^{n-1} e^{-y} dy = \frac{2^{n/2}}{\sqrt{\pi}}\Gamma\left(\frac{n+1}{2}\right)\\
&= 1 \times 3 \times 5 \times \cdots \times (n-1).
\end{aligned}$$

- For a general normal random variable $X \sim \sigma Z + \mu$,

$$\begin{aligned}
E[X] &= \mu\\
\text{Var}(X) &= \sigma^2\\
M_X &= \exp\{t\mu + \sigma^2 t^2/2\}.
\end{aligned}$$

The MGF result follows from a useful general property:

Theorem

If a and b are constants, then

$$M_{aX+b}(t) = e^{tb} M_X(ta)$$

Proof.

Using the definition of the MGF,

$$\begin{aligned} M_{aX+b}(t) &= E\left[e^{t(aX+b)}\right] \\ &= E\left[e^{taX} e^{tb}\right] \\ &= e^{tb} E\left[e^{(ta)X}\right] \\ &= e^{tb} M_X(ta) \end{aligned}$$

□

- The normal distribution is very important as a model and as an approximation.
- For a Binomial(n, p) random variable X, if n is large then X is approximately
$$N(np, np(1-p))$$
- Similar results hold for
  - hypergeometric
  - Poisson
  - gamma
  - negative binomial

Beta Distribution

- X has a Beta(α, β) distribution if it has density

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

  for some positive constants α and β.
- The Beta(1,1) distribution is the uniform distribution on $[0, 1]$.
- Some plots are produced by

```
plot(function(x) dbeta(x, .5, .5), 0, 1)
plot(function(x) dbeta(x, 1, 1), 0, 1)
plot(function(x) dbeta(x, 2, 2), 0, 1)
plot(function(x) dbeta(x, 5, 5), 0, 1)
```

- The normalizing constant of the Beta distribution is the *Beta function*

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

  for $\alpha, \beta > 0$.
- The *incomplete Beta function* is

$$B(x; \alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$$

  for $x \in [0, 1]$.
- The CDF of the Beta distribution is the *regularized incomplete Beta function*

$$F(x) = \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)}$$

  for $x \in [0, 1]$.
- Good numerical algorithms are available for computing these.

- The mean of a Beta(α, β) distribution is

$$\begin{aligned} E[X] &= \int_0^1 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_o^1 x^{\alpha}(1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\alpha}{\alpha+\beta}. \end{aligned}$$

- Similar calculations for $E[X^2]$ produce

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

- The MGF is not useful.

Other Continuous Distributions

I will skip:

- Cauchy
- Lognormal
- Double exponential

You should read about these on your own.

Parametric Families

- In developing models for data, it is often useful to fix some basic characteristics of a model but to allow other features to be adjusted to fit the problem.
- Parametric families are collections of distributions/PMFs/PDFs

$$\{f(x|\theta) : \theta \in \Theta\}$$

  where a particular distribution corresponds to a particular choice of the parameter θ.
- Two useful approaches involve constructing
  - location-scale families
  - exponential families

Location-Scale Families

One very natural way to construct a parametric family:

- Start with a "standard" variable

$$X \sim f$$

- Then define

$Y = X + a$	for some a: a location family
$Y = bX$	for some $b > 0$: a scale family
$Y = bX + a$	for some a, b: a location-scale family

This produces families of densities:

$$g(y|a) = f(y - a) \qquad \text{for some } a\text{: location family}$$

$$g(y|b) = \frac{1}{b} f(y/b) \qquad \text{for some } b > 0\text{: scale family}$$

$$g(y|a, b) = \frac{1}{b} f\left(\frac{y-a}{b}\right) \qquad \text{for some } a, b\text{: location-scale family}$$

Examples

- $f \sim$ Uniform$[0, 1]$, $g \sim$ Uniform$[a, a + b]$.
- $f \sim$ Exponential(1), $g \sim$ Exponential$(1/b)$.
- Location-scale produces a *shifted exponential*.
- $f \sim$ Gamma$(\alpha, 1)$, $g(y|\beta) \sim$ Gamma(α, β).
- $f \sim N(0, 1)$, $g(y|\mu, \sigma) \sim N(\mu, \sigma^2)$.

Adding covariates

- The location parameter in a location-scale family can be related to covariate values, e.g.

$$a = \beta_1 x_1 + \cdots + \beta_p x_p = x^T \beta$$

- This leads to a regression model with *errors* from the scale family based on f.
- Non-linear or *semiparametric* forms are also possible.
- For lifetime data and a distribution f on fhe positive half line, the scale parameter is sometimes modeled as

$$\log b = \beta_1 x_1 + \cdots + \beta_p x_p = x^T \beta$$

- The resulting model is an *accelerated failure time* or AFT model.
- A similar formulation can be used to relate the spread around a regression line to covariates.
- Again nonlinear and semiparametric forms are possible.

- Location-scale families are natural from a modeling point of view.
- Sometimes transforming first helps (e.g. taking logs)
- It is possible to construct families with parameters for
  - location
  - scale
  - skewness
  - kurtosis
- The Pearson families are one approach.
- Unfortunately, location-scale models don't always end up being very easy to work with.
- Location and scale families are also of little use for discrete data.

Recap

- Normal distribution
- Beta distributions
- Location-scale families

Exponential Families

Another approach is to create a family of models that is an *exponential family*.

Definition

A family $\{f(x|\theta) : \theta \in \Theta\}$ of PDF's or PMF's is an *exponential family* if you can write it as

$$f(x|\theta) = h(x) \exp\left\{\sum_{i=1}^{k} w_i(\theta) t_i(x)\right\} c(\theta)$$

for some k and some functions h, c, w_i and t_i.

- The functions h and t_i must only depend on x.
- The functions c and w_i must only depend on θ.
- f must be defined for all of $\mathbb{R}$ with this expression.
- Any range restrictions must be incorporated in h.
- This implies that the range of X cannot depend on θ.
- Often we are a bit sloppy about this.
- Many of the families we have seen are exponential families.

Example

- Suppose $X \sim \text{Binomial}(n, p)$, with $x \in \{0, 1, \ldots, n\}$.
- Then $\theta = p$, and

$$\begin{aligned} f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \left(\frac{p}{1-p} \right)^x (1-p)^n \\ &= \binom{n}{x} \exp\{w(p)x\} c(p) \end{aligned}$$

- So

$$\begin{aligned} h(x) &= \binom{n}{x} 1_{\{0,\ldots,n\}}(x) \\ w(p) &= \log\left(\frac{p}{1-p}\right) \\ t(x) &= x \\ c(p) &= (1-p)^n \end{aligned}$$

Examples

$X \sim$ negative binomial, $\theta = p$, $x = 0, 1, \ldots$.

$$\begin{aligned} f_X(x) &= \binom{r+x-1}{x} p^r (1-p)^x \qquad \text{for } x = 0, 1, 2, \ldots \\ &= \binom{r+x-1}{x} p^r (1-p)^x \, 1_{\{0,1,2,\ldots\}}(x) \\ &= \underbrace{\binom{r+x-1}{x} 1_{\{0,1,2,\ldots\}}(x)}_{h(x)} \underbrace{p^r}_{c(p)} \exp\{\underbrace{x}_{t(x)} \underbrace{\log(1-p)}_{w(p)}\} \end{aligned}$$

Examples (continued)

$X \sim \text{Gamma}(\alpha, \beta)$, $\theta = \beta$, α fixed.

$$\begin{aligned} f_X(x) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} 1_{(0,\infty)}(x) \\ &= \underbrace{\frac{1}{\Gamma(\alpha)} \beta^\alpha}_{c(\beta)} \underbrace{x^{\alpha-1} 1_{(0,\infty)}(x)}_{h(x)} \exp\{\underbrace{-x}_{t(x)} \underbrace{1/\beta}_{w(\beta)}\} \end{aligned}$$

This is also an exponential family for $\theta = (\alpha, \beta)$.

Examples (continued)

$X \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$.

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \\ &= \underbrace{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}}}_{c(\theta)} \exp\{-\underbrace{\frac{x^2}{2}}_{t_1(x)} \underbrace{\frac{1}{\sigma^2}}_{w_1(\theta)} + \underbrace{x}_{t_2(x)} \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\theta)}\} \times \underbrace{1}_{h(x)} \end{aligned}$$

- Sometimes we write an exponential family as

$$f(x|\eta) = h(x)c^*(\eta)\exp\left\{\sum_{i=1}^{k}\eta_i t_i(x)\right\}$$

- The η_i are called the *natural parameters*.
- The set

$$\mathcal{N} = \left\{(\eta_1, \dots, \eta_k) : \int_{-\infty}^{\infty} h(x)\exp\left\{\sum_{i=1}^{k}\eta_i t_i(x)\right\} dx < \infty\right\}$$

is called the *natural parameter space* of the family.

- The mapping
$$(w_1(\theta), \ldots, w_k(\theta)) : \Theta \to \mathcal{N}$$
is a mapping into the natural parameter space.
- Sometimes this mapping is one-to-one and onto.
- Sometimes it may be lower-dimensional.
- For example, $N(\theta, \theta^2)$ is a lower-dimensional mapping.
- This is called a *curved exponential family*.

Adding covariates

- Exponential families are often used as the basis for regression models.
- One (or more) of the parameters is written as a function of covariates.
- For binomial data, a common staring point is

$$\log\left(\frac{p}{1-p}\right) = \beta_1 x_1 + \cdots + \beta_p x_p = x^T\beta$$

- For Poisson data, a common approach is

$$\log\lambda = \beta_1 x_1 + \cdots + \beta_p x_p = x^T\beta$$

- These are examples of *generalized linear models*

- Exponential family structure has a number of advantages:
  - Some expectations can be computed by differentiation.
  - Sufficient statistics are easily identified.
  - Likelihood functions have many nice properties.
- When designing a new model for a complex situation it is a good idea to try to create an exponential family.
- Examples of situations where exponential families have been used are:
  - Models for the distribution of points in the plane.
  - Models for phylogenetic trees in pedigree analysis.
  - Models for social networks in network data analysis.

Multiple Random Variables

- Often we record more than one measurement for an experiment.
  - Roll three dice, record X_1, X_2, X_3
  - Select person at random, record height H and weight W.
- Each is a random variable — a (measurable) real-valued function on a sample space.
- Since each individual variable of $X_1, \dots, X_n$ is a random variable we can study it as before.

- We only need to consider two random variables to see some new issues in looking at several variables jointly.
- We might want to compute

$$P(H > 6', W < 100 lbs)$$
$$P(X_1 + X_2 + X_3 = 8)$$

- In general, we might want to compute

$$P((X_1, \ldots, X_n) \in A)$$

for $A \subset \mathbb{R}^n$ (a Borel set).

Multiple Discrete Random Variables

- Suppose $X_1, \dots, X_n$ are all discrete.
- Then $(X_1, \dots, X_n)$ has finitely or countably many values.
- The *joint PMF* of $X_1, \dots, X_n$ is defined as

$$f(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

- We can use the joint PMF to find any probability,

$$P((X_1, \dots, X_n) \in A) = \sum_{(x_1, \dots, x_n) \in A} f(x_1, \dots, x_n)$$

- The joint PMF must satisfy

$$f(x_1, \dots, x_n) \geq 0$$
$$\sum f(x_1, \dots, x_n) = 1$$

- Any function with these two properties is a joint PMF.