



# Some Common Distributions

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng  
Johns Hopkins Bloomberg School of Public Health

# The Bernoulli distribution

- The Bernoulli distribution arises as the result of a binary outcome
- Bernoulli random variables take (only) the values 1 and 0 with probabilities of (say)  $p$  and  $1 - p$  respectively
- The PMF for a Bernoulli random variable  $X$  is

$$P(X = x) = p^x(1 - p)^{1-x}$$

- The mean of a Bernoulli random variable is  $p$  and the variance is  $p(1 - p)$
- If we let  $X$  be a Bernoulli random variable, it is typical to call  $X = 1$  as a "success" and  $X = 0$  as a "failure"

# iid Bernoulli trials

- If several iid Bernoulli observations, say  $x_1, \dots, x_n$ , are observed the likelihood is

e.g. for  $n=4$ :  
 $p(1, 1, 0, 0) = p(0, 1, 0, 1) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}$   $\Rightarrow$  just the product of the mass function, because of independence  
(Reihenfolge egal)

- Notice that the likelihood depends only on the sum of the  $x_i$
- Because  $n$  is fixed and assumed known, this implies that the sample proportion  $\sum_i x_i / n$  contains all of the relevant information about  $p$
- We can maximize the Bernoulli likelihood over  $p$  to obtain that  $\hat{p} = \sum_i x_i / n$  is the maximum likelihood estimator for  $p$

D.h. wenn man  $p$  von 0..1 variiert, wird die Bernoulli likelihood maximal bei  $\hat{p} = \sum_i (x_i) / n$  ?  
Das hiesse zB

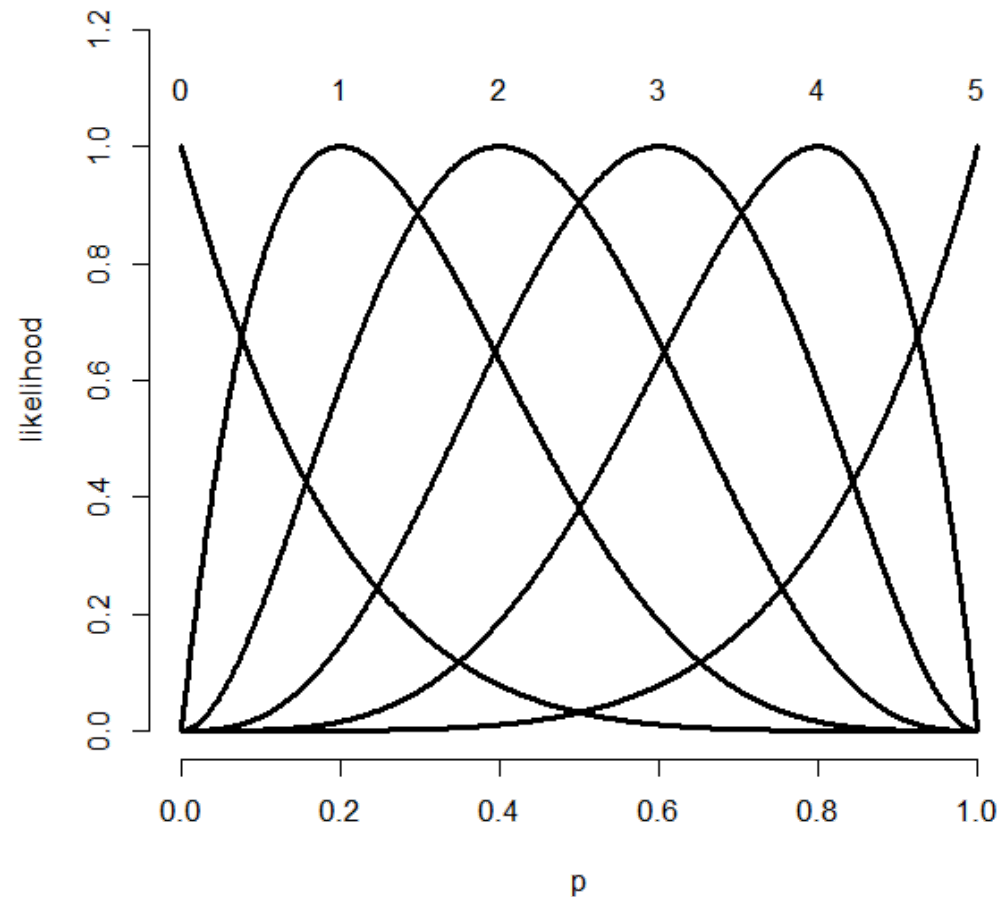
fuer  $p(1,1,0,0)$  ist max Bernoulli likelihood bei  $\hat{p} = 2/4 = 0.5$

fuer  $p(1,0,1,1)$  ist max Bernoulli likelihood bei  $\hat{p} = 3/4 = 0.75$

# Plotting all possible likelihoods for a small n

```
n <- 5
pvals <- seq(0, 1, length = 1000)
plot(c(0, 1), c(0, 1.2), type = "n", frame = FALSE, xlab = "p", ylab = "likelihood")
text((0 : n) / n, 1.1, as.character(0 : n))
sapply(0 : n, function(x) {
  phat <- x / n
  if (x == 0) lines(pvals, ( (1 - pvals) / (1 - phat) )^(n-x), lwd = 3)
  else if (x == n) lines(pvals, (pvals / phat) ^ x, lwd = 3)
  else lines(pvals, (pvals / phat) ^ x * ( (1 - pvals) / (1 - phat) ) ^ (n-x), lwd = 3)
})
title(paste("Likelihoods for n = ", n))
```

Likelihoods for  $n = 5$



zB fuer 1 success in 5 trials: max likelihood bei  $1/5=0.2$   
fuer 2 success in 5 trials: max likelihood bei  $2/5=0.4$   
etc

# Binomial trials

- The binomial random variables are obtained as the sum of iid Bernoulli trials
- In specific, let  $X_1, \dots, X_n$  be iid Bernoulli( $p$ ); then  $X = \sum_{i=1}^n X_i$  is a binomial random variable
- The binomial mass function is

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

for  $x = 0, \dots, n$

| this part is the Bernoulli mass function

|

| and this part gives how many possibilities there are  
| to get x successes from n trials

Since all the same probabilities, we need to multiply  
this number by that probability, which is the second  
part of the formula.

Bernoulli:  $(1,0,1,0) = (1,1,0,0) = (0,1,1,0)$  etc

Binomial: Just the sum of the Bernoullis: hier = 2

# Choose

- Recall that the notation

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

(read " $n$  choose  $x$ ") counts the number of ways of selecting  $x$  items out of  $n$  without replacement disregarding the order of the items

$$\binom{n}{0} = \binom{n}{n} = 1$$

# Example justification of the binomial likelihood

- Consider the probability of getting 6 heads out of 10 coin flips from a coin with success probability  $p$
- The probability of getting 6 heads and 4 tails in any specific order is

$$p^6(1-p)^4 \quad = \text{Bernoulli prob}$$

- There are

$$\binom{10}{6}$$

possible orders of 6 heads and 4 tails

und diese beiden Teile multipliziert gibt die Binomiale prob.



# Example

- Suppose a friend has 8 children (oh my!), 7 of which are girls and none are twins
- If each gender has an independent 50% probability for each birth, what's the probability of getting 7 or more girls out of 8 births?

$$\binom{8}{7} \cdot 0.5^7 (1 - 0.5)^1 + \binom{8}{8} \cdot 0.5^8 (1 - 0.5)^0 \approx 0.04$$

```
choose(8, 7) * .5 ^ 7 + choose(8, 8) * .5 ^ 8
```

```
[1] 0.03516
```

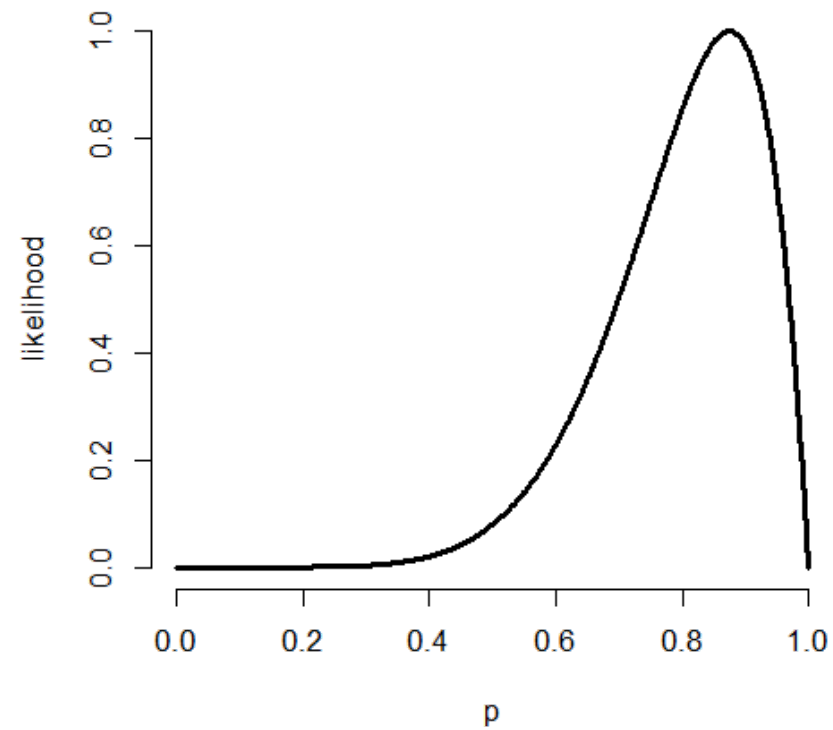
```
pbinom(6, size = 8, prob = .5, lower.tail = FALSE)
```

False => zaehlt alle > 6, also 7 und 8

True => zaehlt alle bis und mit 6

```
[1] 0.03516
```

```
plot(pvals, dbinom(7, 8, pvals) / dbinom(7, 8, 7/8) ,  
     lwd = 3, frame = FALSE, type = "l", xlab = "p", ylab = "likelihood")
```



# The normal distribution

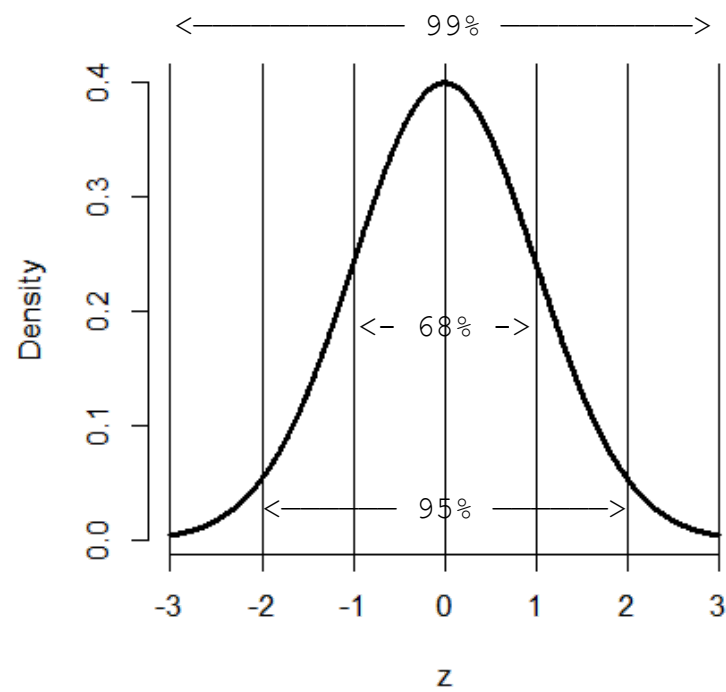
- A random variable is said to follow a normal or Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$  if the associated density is

$$(2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}$$

If  $X$  a RV with this density then  $E[X] = \mu$  and  $Var(X) = \sigma^2$

- We write  $X \sim N(\mu, \sigma^2)$
- When  $\mu = 0$  and  $\sigma = 1$  the resulting distribution is called the standard normal distribution
- The standard normal density function is labeled  $\phi$
- Standard normal RVs are often labeled  $Z$

```
zvals <- seq(-3, 3, length = 1000)
plot(zvals, dnorm(zvals),
     type = "l", lwd = 3, frame = FALSE, xlab = "z", ylab = "Density")
sapply(-3 : 3, function(k) abline(v = k))
```



```
[[1]]
```

```
NULL
```

```
[[2]]
```

```
NULL
```

# Facts about the normal density

- If  $X \sim N(\mu, \sigma^2)$  the  $Z = \frac{X - \mu}{\sigma}$  is standard normal
- If  $Z$  is standard normal

$$X = \mu + \sigma Z \sim N(\mu, \sigma^2)$$

- The non-standard normal density is

$$\phi\{(x - \mu)/\sigma\}/\sigma$$

Weil  $E[Z] = E[(X - \mu)/\sigma] = 1/\sigma * E[X - \mu] = 1/\sigma * (E[X] - \mu) = 0$  hx  
Und  $\text{var}(Z) = \text{var}(X - \mu)/\sigma^2 = 1/\sigma^2 * \text{var}(X - \mu)$  //  $-\mu$  ändert nichts an der Varianz, also:  
 $= 1/\sigma^2 * \text{var}(X)$   
 $= 1/\sigma^2 * \sigma^2 = 1$  hx

# More facts about the normal density

1. Approximately 68%, 95% and 99% of the normal density lies within 1, 2 and 3 standard deviations from the mean, respectively
2.  $-1.28$ ,  $-1.645$ ,  $-1.96$  and  $-2.33$  are the  $10^{th}$ ,  $5^{th}$ ,  $2.5^{th}$  and  $1^{st}$  percentiles of the standard normal distribution respectively
3. By symmetry,  $1.28$ ,  $1.645$ ,  $1.96$  and  $2.33$  are the  $90^{th}$ ,  $95^{th}$ ,  $97.5^{th}$  and  $99^{th}$  percentiles of the standard normal distribution respectively

# Question

- What is the 95<sup>th</sup> percentile of a  $N(\mu, \sigma^2)$  distribution?
  - Quick answer in R `qnorm(.95, mean = mu, sd = sd)`
- We want the point  $x_0$  so that  $P(X \leq x_0) = .95$

$$\begin{aligned} P(X \leq x_0) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{x_0 - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x_0 - \mu}{\sigma}\right) = .95 \end{aligned}$$

- Therefore

$$\frac{x_0 - \mu}{\sigma} = 1.645$$

or  $x_0 = \mu + \sigma 1.645$

- In general  $x_0 = \mu + \sigma z_0$  where  $z_0$  is the appropriate standard normal quantile

# Question

- What is the probability that a  $N(\mu, \sigma^2)$  RV is 2 standard deviations above the mean?
- We want to know

$$\begin{aligned}P(X > \mu + 2\sigma) &= P\left(\frac{X - \mu}{\sigma} > \frac{\mu + 2\sigma - \mu}{\sigma}\right) \\&= P(Z \geq 2) \\&\approx 2.5\%\end{aligned}$$



# Other properties

- The normal distribution is symmetric and peaked about its mean (therefore the mean, median and mode are all equal)
- A constant times a normally distributed random variable is also normally distributed (what is the mean and variance?)
- Sums of normally distributed random variables are again normally distributed even if the variables are dependent (what is the mean and variance?)
- Sample means of normally distributed random variables are again normally distributed (with what mean and variance?)
- The square of a standard normal random variable follows what is called chi-squared distribution
- The exponent of a normally distributed random variables follows what is called the log-normal distribution
- As we will see later, many random variables, properly normalized, limit to a normal distribution

Darum ist die Normalverteilung so wichtig!

# Final thoughts on normal likelihoods

- The MLE for  $\mu$  is  $\bar{X}$ . MLE: Maximum Likelihood Estimator
- The MLE for  $\sigma^2$  is

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

(Which is the biased version of the sample variance.)

- The MLE of  $\sigma$  is simply the square root of this estimate

# The Poisson distribution

- Used to model counts
- The Poisson mass function is

$$P(X = x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

for  $x = 0, 1, \dots$

- The mean of this distribution is  $\lambda$
- The variance of this distribution is  $\lambda$  also  $E[X] = \text{var}(X) \quad !$
- Notice that  $x$  ranges from 0 to  $\infty$  zB, wie viele Leute kommen ans Konzert?  
-> also potentiell unendlich viele moeglich.

# Some uses for the Poisson distribution

- Modeling event/time data
- Modeling radioactive decay
- Modeling survival data    Modeling time until some event (classically death)
- Modeling unbounded count data    - How many people show up at the bus stop?
- Modeling contingency tables
- Approximating binomials when  $n$  is large and  $p$  is small

eg disease that is very rare and a large population

# Poisson derivation

- $\lambda$  is the mean number of events per unit time
- Let  $h$  be very small
- Suppose we assume that
  - Prob. of an event in an interval of length  $h$  is  $\lambda h$  while the prob. of more than one event is negligible
  - Whether or not an event occurs in one small interval does not impact whether or not an event occurs in another small interval then, the number of events per unit time is Poisson with mean  $\lambda$

# Rates and Poisson random variables

- Poisson random variables are used to model rates
- $X \sim \text{Poisson}(\lambda t)$  where
  - $\lambda = E[X/t]$  is the expected count per unit of time
  - $t$  is the total monitoring time

# Poisson approximation to the binomial

- When  $n$  is large and  $p$  is small the Poisson distribution is an accurate approximation to the binomial distribution
  - Notation
    - $\lambda = np$
    - $X \sim \text{Binomial}(n, p)$  ,  $\lambda = np$  and
    - $n$  gets large
    - $p$  gets small
    - $\lambda$  stays constant
- Poisson approx to binomial is /very/ common. To the point where ppl don't even mention they're using an approximation.

# Example

The number of people that show up at a bus stop is Poisson with a mean of 2.5 per hour.

If watching the bus stop for 4 hours, what is the probability that 3 or fewer people show up for the whole time?

```
ppois(3, lambda = 2.5 * 4)
```

```
[1] 0.01034
```



# Example, Poisson approximation to the binomial

We flip a coin with success probability 0.01 five hundred times.

What's the probability of 2 or fewer successes?

```
pbinom(2, size = 500, prob = .01)
```

```
[1] 0.1234
```

```
ppois(2, lambda=500 * .01)
```

```
[1] 0.1247
```