Mathematical Statistics

Anna Janicka

Lecture XV, 9.06.2022

BAYESIAN STATISTICS

& SOME CONCLUDING REMARKS

Plan for Today

1. Bayesian Statistics
   - prior and a posterior distributions
   - Bayesian estimation:
     - Bayesian Most Probable Estimator (BMP)
     - Bayes Estimator
2. Caution!

Bayesian Statistics vs. traditional statistics

Frequentist: unknown parameters are given (fixed), observed data are random

Bayesian: observed data are given (fixed), **parameters are random**

Bayesian Statistics

Our knowledge about the unknown parameters is described by means of probability distributions, and additional knowledge may affect our description.

Knowledge:

- general
- specific

Example: coin toss

Bayesian Model

- $X_1, \ldots, X_n$ come from distribution P_θ, with density $f_\theta(\mathbf{x})$ – conditional density given a specific value of θ (likelihood function).
- $\mathscr{P}$ – family of probability distributions P_θ, indexed by the parameter $\theta \in \Theta$
- General knowledge: distribution Π over the parameter space Θ, given by $\pi(\theta)$ – the so-called **prior** distribution of θ,

$$\theta \sim \Pi$$

Bayesian Model – cont.

Additional knowledge (specific, contextual): based on observation. We have a joint distribution of observations and θ:

$$f(x_1, x_2, \dots, x_n, \theta) = f(x_1, x_2, \dots, x_n|\theta)\pi(\theta)$$

on this basis we can derive the conditional distribution of θ (given the observed data)

$$\pi(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta)\pi(\theta)}{m(x_1, \dots, x_n)},$$

where

$$m(x_1, \dots, x_n) = \int_\Theta f(x_1, \dots, x_n|\theta)\pi(\theta)d\theta$$

is a marginal distribution for the obs.

Bayesian Model – a posteriori distribution

$\pi(\theta|x_1,\ldots,x_n)$ is called the **posterior** distribution, denoted Π_x

The posterior distribution reflects all knowledge: general (initial) and specific (based on the observed data).

Grounds for Bayesian inference and modeling

Prior and posterior distributions: examples

Let $X_1, \ldots, X_n$ be IID r.v. from a 0-1 distr. with prob. of success θ; let

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

for $\theta \in (0,1)$

Beta(α,β) distr with mean $= \alpha/(\alpha+\beta)$

where $$B(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

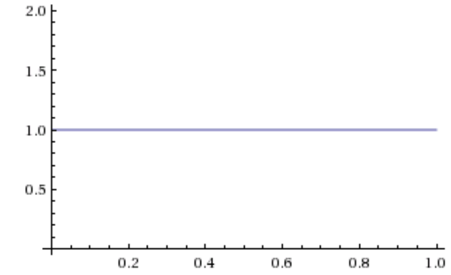
and $$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}\exp(-u)du = (\alpha-1)\Gamma(\alpha-1)$$

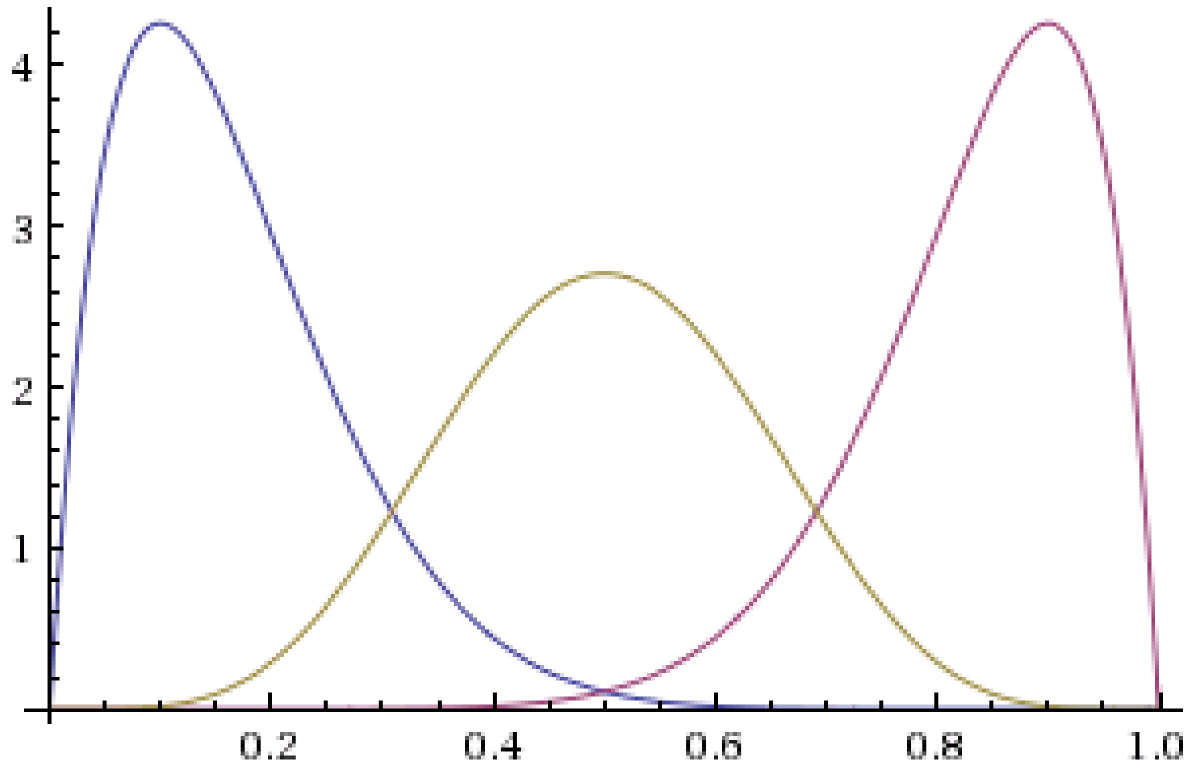
then the posterior distribution:

$$\text{Beta}\left(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta\right)$$

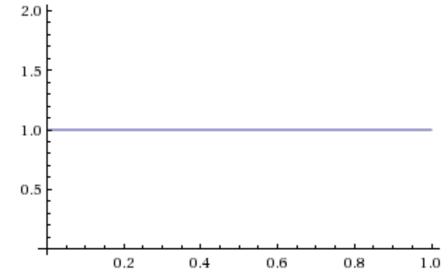
conjugate prior for Bernoulli distr.

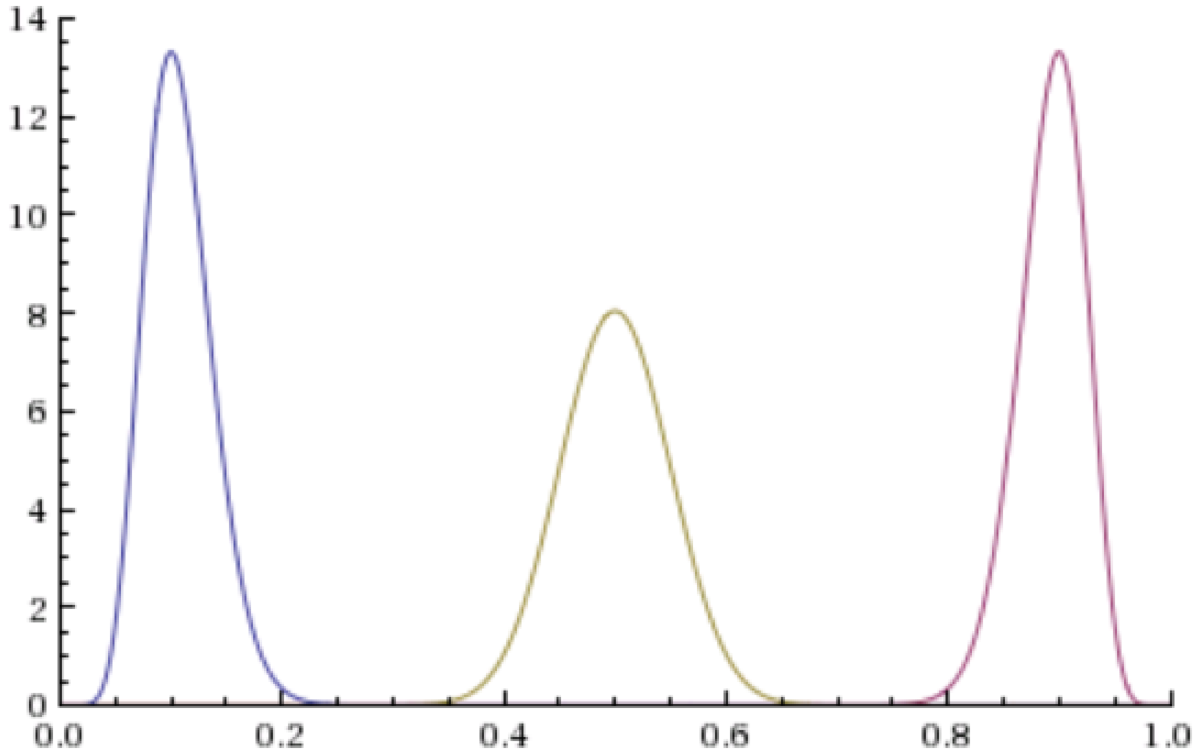
For a Beta (1,1) prior and data: n=10 and 1, 5, 9 successes



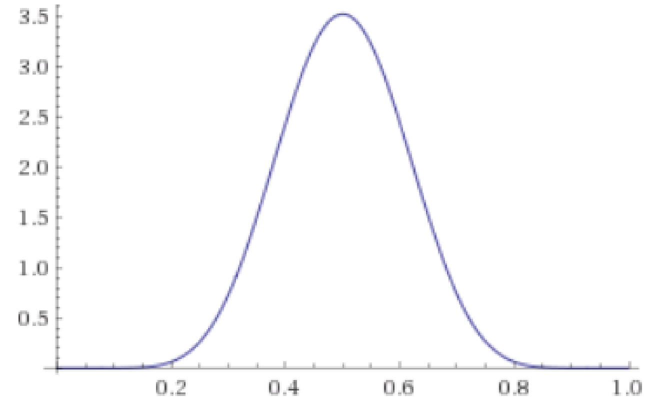


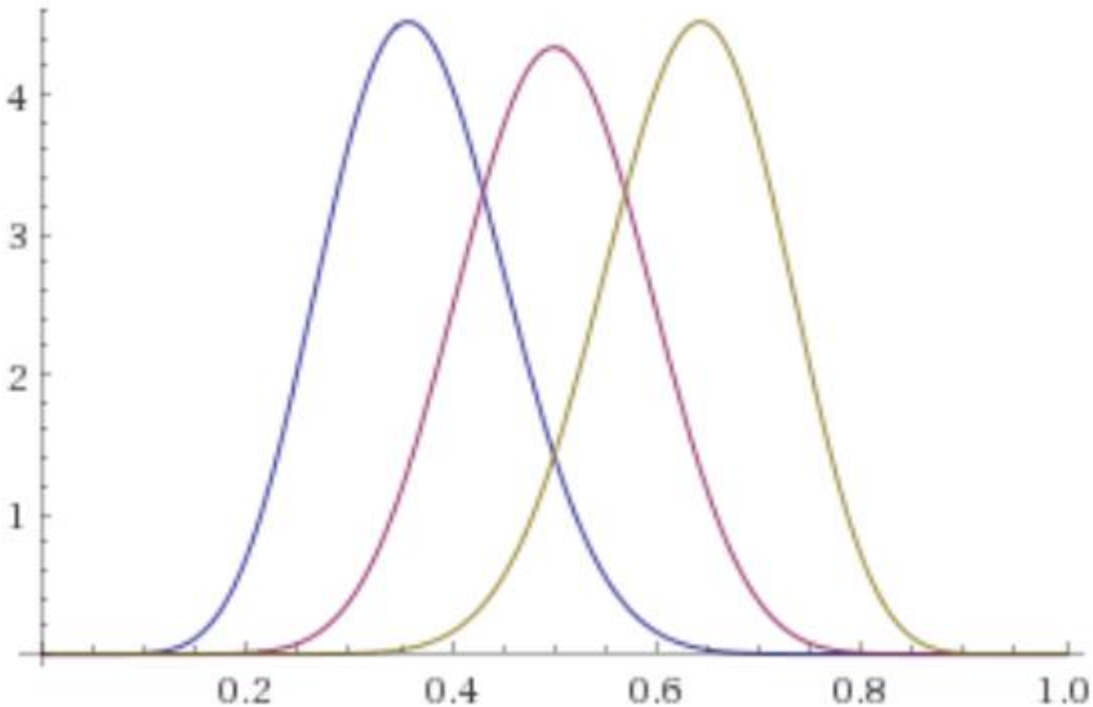
For a Beta (1,1) prior and data: n=100 and 10, 50, 90 successes



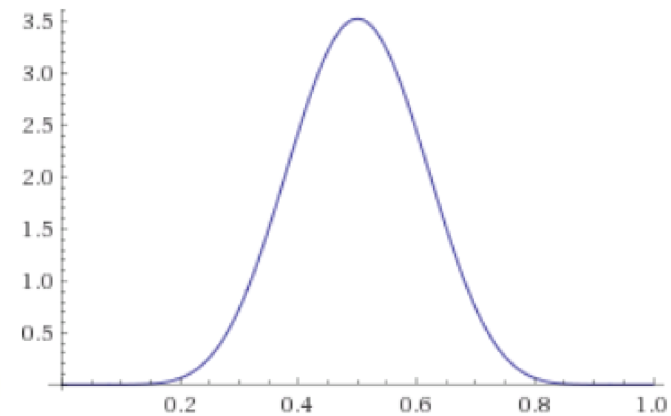


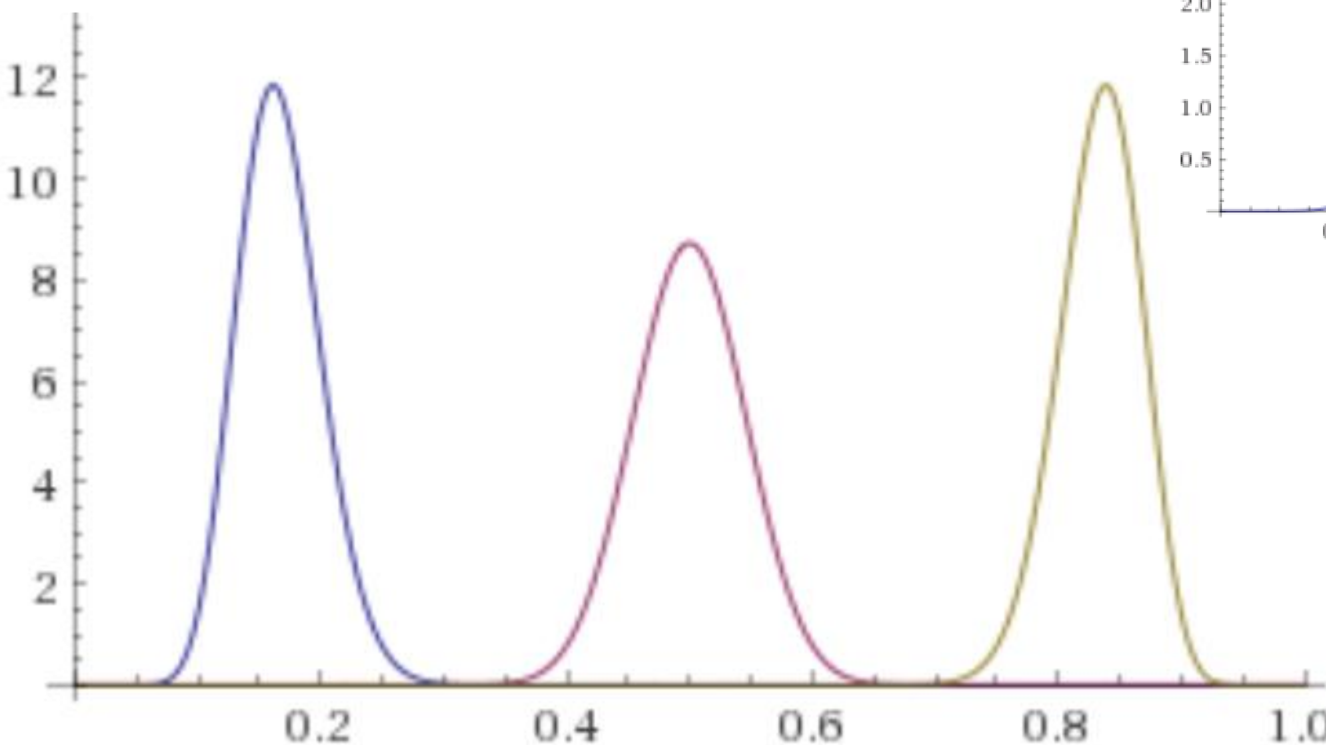
For a Beta (10,10) prior and data: n=10 and 1, 5, 9 successes



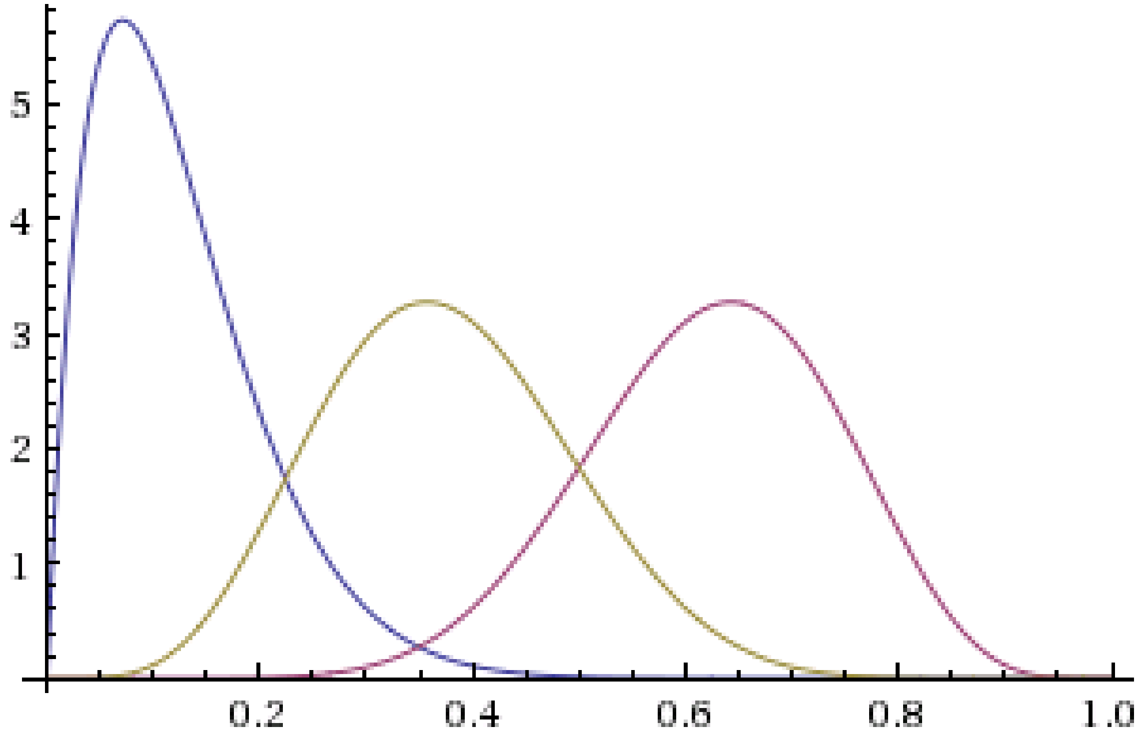


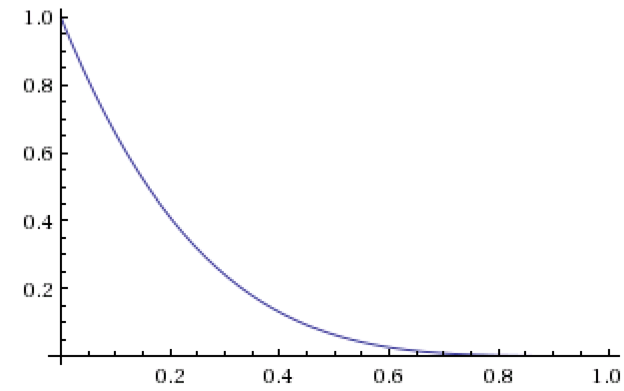
For a Beta (10,10) prior and data: n=100 and 10, 50, 90 successes



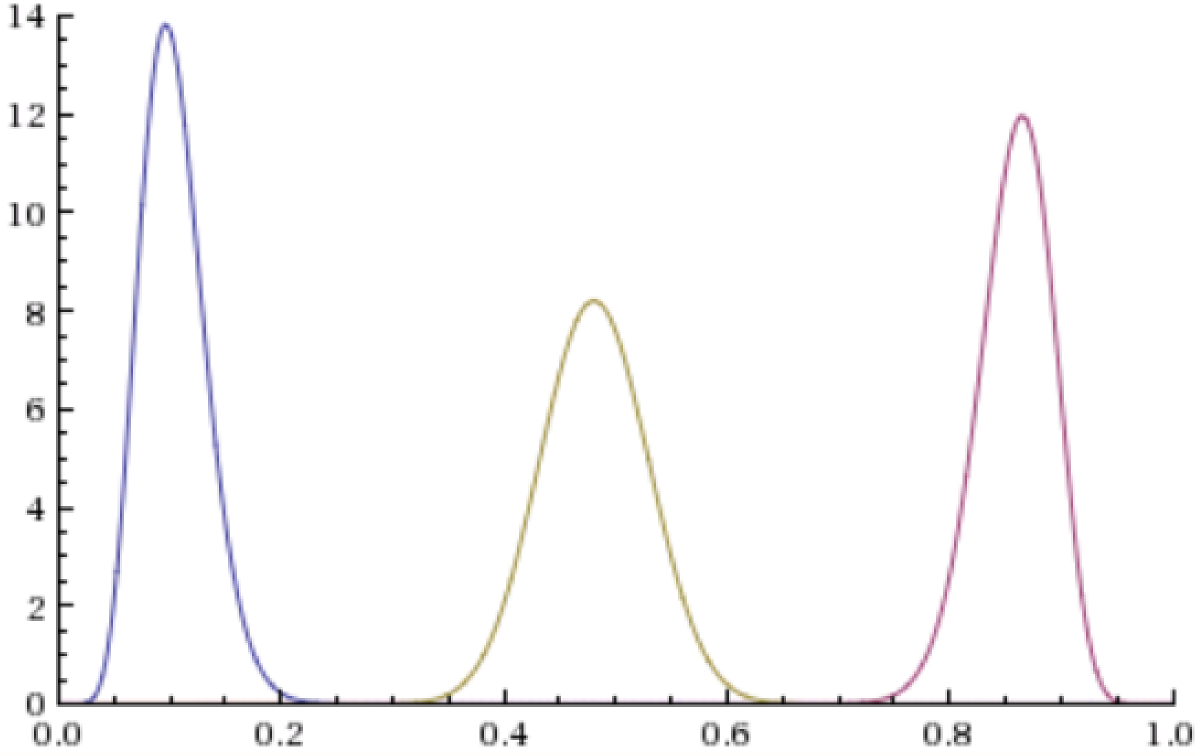


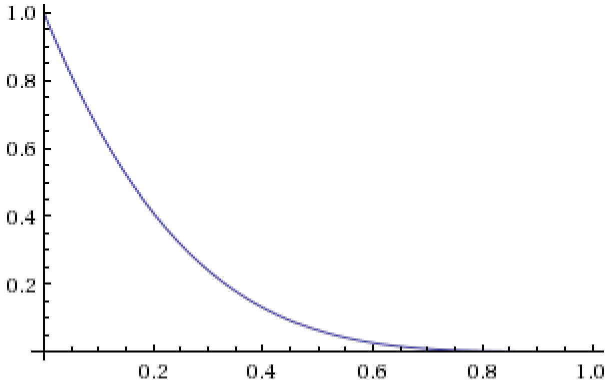
For a Beta (1,5) prior and data: n=10 and 1, 5, 9 successes





For a Beta (1,5) prior and data: n=100 and 10, 50, 90 successes





Bayesian Statistics

Based on the Bayes approach, we can

- find estimates
- find an equivalent of confidence intervals
- verify hypotheses

- make predictions

Bayesian Most Probabale (BMP) estimate

Similar to ML estimation: the argument which maximizes the posterior distribution:

$$\pi(\hat{\theta}_{BMP}|x_1,\ldots,x_n) = \max_\theta \pi\,(\theta|x_1,\ldots,x_n)$$

i.e.

$$BMP(\theta) = \hat{\theta}_{BMP} = \operatorname{argmax}_\theta \pi\,(\theta|x_1,\ldots,x_n)$$

BMP: example

1. Let $X_1, ..., X_n$ be IID r.v. from a Bernoulli distr. with prob. of success θ; for $\theta \in (0,1)$

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

We know the posterior distribution:

$$\mathrm{Beta}(\sum_{i=1}^{n} x_i + \alpha, n - \sum_{i=1}^{n} x_i + \beta)$$

Beta(α,β) distr; the mode of this distr = (α-1)/(α+ β-2) for α>1, β>1

we have max for

$$BMP(\theta) = \frac{\sum_{i=1}^{n} x_i + \alpha - 1}{n + \beta + \alpha - 2}$$

i.e. for 5 successes in 10 trials for a prior U(0,1) (i.e. Beta(1,1) distr.), we have *BMP*(θ)=5/10 = ½

and for 9 successes in 10 trials for the same prior distr., we have *BMP*(θ)=9/10

Bayes Estimator

An estimation rule which minimizes the posterior expected value of a loss function

This is equivalent to using other (than the mode) characteristics of the posterior distribution to find an estimator:

- the mean
- the median

Caution!

1. Tests should be designed BEFORE we start examining the data
2. The only way to increase power and improve significance level simultaneously is by collecting more observations (frequently not possible if we work on existing data). Attn to two populations!
3. Significant p-value does not mean effect is important/sizeable.
4. P-values in repeated samples

P-values in repeated samples

We examine if a new training has effect. The null hypothesis is that the training has no effect, and the alternative hypothesis is that it has effect. We use a 5% significance level for the test.

- A randomly selected school has completed this training, and after completion the statistical test returns a *P*-value equal to 4%.
- 25 different schools have completed this training. At one of the schools the test returned a *P*-value of 4%.

Order of actions

- What is the alternative hypothesis (the one we want to prove)?
- What is the null hypothesis (the one we want to disprove)?
- Which test statistic should we use?
- When should we reject the null hypothesis?

BEFORE WE START EXAMINING THE DATA (preferably before the experiment design)

Null and alternative hypotheses: examples

- A firm claims that more than 50% of the population prefer their new product. We ask n randomly selected people if they prefer the new product and we register X, the number of people in the sample who answer yes. We believe that the company may be right, and wish to execute a test where it will be possible to conclude that the firm probably is right.
- A firm claims that at most 10% of the customers are dissatisfied with the items they have bought from the firm. We ask n randomly selected customers if they are dissatisfied and register X, the number of customers who are dissatisfied. We believe the firm is mistaken and want to execute a test where it is possible to conclude that the firm probably is mistaken.

When is the alternative one-sided and when is it two-sided?

- We examine whether some special form of training leads to improved production. We measure production in terms of an unknown parameter which increases when production improves
- We examine whether some form of new security measure affects production.

Attn: Paired vs unpaired

A factory can use two different methods of production. We make 10 independent observations of the production, 5 using method 1 and 5 using method 2. Method 1 gave the results:

4.7; 3.5; 3.3; 4.2; 3.6;

while the corresponding numbers for method 2 were

3.2; 4.2; 3.3; 3.9; 3.0.

Assuming normality and equality of variances, the t-test for this sample is T=0.99 with critical value 2.306.

If we look at the results ordering 5 workers, we have:

4.7; 3.5; 3.3; 4.2; 3.6 and 4.2; 3.2; 3.0; 3.9; 3.3

→ Paired test with different outcome!

Attn: Independence of observations

We observe stock prices of a company, we want to verify if there is an increasing trend.

Is it reasonable to assume that observations are independent?

X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
100	92	96	117	120	126	149	152	176	196	184

We use a transformation $Y_i = \ln(\frac{X_i}{X_{i-1}})$

Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
-0.083	0.043	0.198	0.025	0.049	0.168	0.02	0.147	0.108	-0.063

Source: Jan Ubøe, Introductory Statistics for Business and Economics

Attn: Extremes

- A company has 10 machines, all units produce on average m items per day with a standard deviation of 5 items.
  - Assuming normality, the critical value for a 5% significance level test for m=100 against m<100 is approximately 92
  - Assuming normality, and m=100, what is the probability that at least one plant produces less than 92?
  - What is the critical value for testing that at least one out of the ten units produces less than 100 by looking at the minimum production value?

Final quiz

https://forms.gle/gaKZgozdGiEjWn768