Mathematical Statistics

Anna Janicka

Lecture XIII, 28.05.2018

NON-PARAMETRIC TESTS

BAYESIAN STATISTICS – INTRO

Plan for Today

1. Goodness-of-fit tests
   - Kolmogorov test
   - Kolmogorov-Smirnov (two samples)
   - Kolmogorov-Lilliefors
   - chi-squared goodness-of-fit
2. Tests of independence
   - chi-squared test
3. <u>Bayesian Statistics</u>
   - a priori and a posteriori distributions
   - Bayesian estimation:
     - Maximum a posteriori probability (MAP)
     - Bayes Estimator

Non-parametric tests

- we check whether a random variable fits a given distribution (goodness-of-fit tests).
- we check whether random variables have the same distribution
- we check whether variables/characteristics are independet (test of independence)

Kolmogorov goodness-of-fit test

Model: $X_1, X_2, \ldots, X_n$ are an IID sample from distribution with CDF F.

H_0: $F = F_0$ (F_0 specified)

H_1: $\neg H_0$ (i.e. the CDF is different)

If F_0 is continuous, we use the statistic

$$D_n = \sup_{t \in R} |F_n(t) - F_0(t)| = \max\{D_n^+, D_n^-\}$$

where

$$D_n^+ = \max_{i=1,\ldots,n} \left| \frac{i}{n} - F_0(x_{i:n}) \right|, \quad D_n^- = \max_{i=1,\ldots,n} \left| F_0(x_{i:n}) - \frac{i-1}{n} \right|$$

and $F_n(t)$ – n-th empirical CDF

Kolmogorov goodness-of-fit test – cont.

The test: we reject H_0 when:

$$D_n > c(\alpha, n)$$

for a critical value $c(\alpha, n)$.

Theorem. If H_0 is true, the distribution of D_n does not depend on F_0.

Problem: This distribution needs tables, for each different *n.*

Theorem. In the limit

$$P(\sqrt{n}D_n \leq d) \xrightarrow[n \to \infty]{} K(d) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2d^2}$$

the approximation may be used for $n \geq 100$

Kolmogorov goodness-of-fit test – cont. (2)

Tables of the asymptotic distribution $K(d)$

$1-\alpha$	0.8	0.9	0.95	0.99
quantile of $K(d)$	1,07	1,22	1,36	1,63
$c(n, \alpha)$ for $n \geq 100$	$1{,}07/\sqrt{n}$	$1{,}22/\sqrt{n}$	$1{,}36/\sqrt{n}$	$1{,}63/\sqrt{n}$

Kolmogorov goodness-of-fit test – example

Does the sample

0.4085	0.5267	0.3751	0.8329	0.0846
0.8306	0.6264	0.3086	0.3662	0.7952

come from a uniform distribution U(0,1)?

Source: W. Niemiro

Kolmogorov goodness-of-fit test – example cont.

$X_{i:10}$	(i-1)/10	i/10	i/10 - F($X_{i:10}$)	F($X_{i:10}$)-i/10
0.0846	0	0.1	0.0154	0.0846
0.3086	0.1	0.2	-0.1086	**0.2086**
0.3662	0.2	0.3	-0.0662	0.1662
0.3751	0.3	0.4	0.0249	0.0751
0.4085	0.4	0.5	0.0915	0.0085
0.5267	0.5	0.6	0.0733	0.0267
0.6264	0.6	0.7	0.0736	0.0264
0.7952	0.7	0.8	0.0048	0.0952
0.8306	0.8	0.9	0.0694	0.0306
0.8329	0.9	1	**0.1671**	-0.0671

$D_n = 0.2086 \qquad c(10; 0.9) = 0.369$

→ no grounds to reject the null hypothesis that the distribution is uniform

Kolmogorov-Smirnov test of equality of distributions

Model: $X_1, X_2, ..., X_n$ are an IID sample from a distribution with CDF F, $Y_1, Y_2, ..., Y_m$ are an IID sample from a distribution with CDF G.

H_0: $F = G$

H_1: $\neg H_0$ (i.e. the CDF functions/distributions differ)

If F (and G) is continuous, we test with

$$D_{n,m} = \sup_{t \in R} |F_n(t) - G_m(t)|$$

where $F_n(t)$ – n-th empirical CDF for the first sample, and $G_m(t)$ – m-th empirical CDF for the second sample

Kolmogorov-Smirnov test of equality of distributions – cont.

The test: we reject H_0 if:

$$D_{n,m} > c(\alpha, n, m)$$

for a critical value $c(\alpha, n, m)$.

Theorem. If H_0 is true, the distribution of $D_{n,m}$ does not depend on F (or G).

Theorem. In the limit

$$P\left(\sqrt{\tfrac{nm}{n+m}} D_{n,m} \le d\right) \xrightarrow[n\to\infty, m\to\infty]{} K(d) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 d^2}$$

the approximation is OK for $n,m \ge 100$

Kolmogorov-Lilliefors goodness-of-fit test

Model: X_1, X_2, ..., X_n are an IID sample from a distribution with CDF F.

H_0: F is a CDF of a normal distribution

(**with unknown parameters**)

H_1: $\neg H_0$ (i.e. the distribution is not normal)

We test with

$$D_n = \max\{D_n^+, D_n^-\}$$

where

$$D_n^+ = \max_{i=1,\dots,n}\left|\frac{i}{n} - z_i\right|,\ D_n^- = \max_{i=1,\dots,n}\left|z_i - \frac{i-1}{n}\right|$$

and

$$z_i = \Phi\left(\frac{X_{i:n} - \overline{X}}{S}\right)$$

$$\overline{X} = \tfrac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \tfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Kolmogorov-Lilliefors goodness-of-fit test – cont.

The test: we reject H_0 if:

$$D_n > D_n(\alpha)$$

for a critical value $D_n(\alpha)$.

Theorem. If H_0 is true, the distribution of D_n does not depend on the parameters of the normal distribution.

Problem: we need tables and do not know the analytical form of this distribution...

Used for small samples ($n \leq 30$), when it performs better than the chi-squared test

Kolmogorov-Lilliefors goodness-of-fit test – critical values

Sample Size N	Level of Significance for $D = \text{Max}\left|F^*(X) - S_N(X)\right|$				
	.20	.15	.10	.05	.01
4	.300	.319	.352	.381	.417
5	.285	.299	.315	.337	.405
6	.265	.277	.294	.319	.364
7	.247	.258	.276	.300	.348
8	.233	.244	.261	.285	.331
9	.223	.233	.249	.271	.311
10	.215	.224	.239	.258	.294
11	.206	.217	.230	.249	.284
12	.199	.212	.223	.242	.275
13	.190	.202	.214	.234	.268
14	.183	.194	.207	.227	.261
15	.177	.187	.201	.220	.257
16	.173	.182	.195	.213	.250
17	.169	.177	.189	.206	.245
18	.166	.173	.184	.200	.239
19	.163	.169	.179	.195	.235
20	.160	.166	.174	.190	.231
25	.149	.153	.165	.180	.203
30	.131	.136	.144	.161	.187
Over 30	$\frac{.736}{\sqrt{N}}$	$\frac{.768}{\sqrt{N}}$	$\frac{.805}{\sqrt{N}}$	$\frac{.886}{\sqrt{N}}$	$\frac{1.031}{\sqrt{N}}$

Source: H. Lilliefors

Chi-squared goodness-of-fit test

Model: $X_1, X_2, \ldots, X_n$ are an IID sample from a discrete distribution with k values $(1, \ldots, k)$.

H_0: the distribution probabilities are equal to

i	1	2	3	...	k
$P(X=i)$	p_1	p_2	p_3	...	p_k

H_1: $\neg H_0$ (i.e. the distribution is different)

value labels

If the results of the experiment are

i	1	2	3	...	k
N_i	N_1	N_2	N_3	...	N_k

where N_i denotes the number of outcomes equal to i: $N_i = \sum_{j=1}^{n} 1_{X_j = i}$

Chi-squared goodness-of-fit test – cont.

General form of the test:

$$\chi^2 = \sum \frac{(\text{observed value - expected value})^2}{\text{expected value}}$$

here:

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$$

Theorem. If H_0 is true, the distribution of the χ^2 statistic converges to a chi-squared distr with k-1 degrees of freedom $\chi^2(k\text{-}1)$ for $n \to \infty$

Procedure: we reject H_0 if $\chi^2 > c$, where $c = \chi^2_{1-\alpha}(k\text{-}1)$ is a quantile of rank 1-α from a chi-squared distr with k-1 degrees of freedom

Chi-squared goodness-of-fit test – example

Is a die symmetric? For a significance level α=0.05

n=150 tosses. Results:

i	1	2	3	4	5	6
N_i	15	27	36	17	26	29

H_0: $(N_1, N_2, N_3, N_4, N_5, N_6)$

$\sim \text{Mult}(150, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$

H_1: $\neg H_0$

$$\chi^2 = \frac{(15-25)^2}{25} + \frac{(27-25)^2}{25} + \frac{(36-25)^2}{25} + \frac{(17-25)^2}{25} + \frac{(26-25)^2}{25} + \frac{(29-25)^2}{25} = 12.24$$

$\chi^2_{1-0.05}(5) \approx 11.7$ $\rightarrow$ **we reject H_0.**

Source: W. Niemiro

Chi-squared goodness-of-fit test – distribution with an unknown parameter.

Model: $X_1, X_2, \ldots, X_n$ are an IID sample from a discrete distribution with k values $(1, \ldots, k)$.

H_0: distribution probabilities are equal to

i	1	2	3	...	k
$P(X=i)$	$p_1(\theta)$	$p_2(\theta)$	$p_3(\theta)$	...	$p_k(\theta)$

where θ is an unknown parameter of dimension d.

H_1: $\neg H_0$ (i.e. the distribution is different)

Chi-squared goodness-of-fit test – distribution with an unknown parameter, cont.

Test statistics are constructed like in the previous case, with the expected values calculated using ML estimators of the parameter θ. Only the number of degrees of freedom changes:

Theorem. If H_0 is true, the distribution of the χ^2 statistic converges to a chi-squared distribution with k-d-1 degrees of freedom $\chi^2(k\text{-}d\text{-}1)$ for $n \rightarrow \infty$

Chi-squared goodness-of-fit test – version for continuous distributions

Kolmogorov tests are better, but the chi-squared test may also be used

Model: $X_1, X_2, ..., X_n$ are an IID sample from a continuous distribution.

H_0: The distribution is given by F

H_1: $\neg H_0$ (i.e. the distribution is different)

It suffices to divide the range of values of the random variable into classes and count the observations. The expected values are known (result from F). Then: the chi-squared test.

Chi-squared goodness-of-fit test – practical notes

- The test should be used for large samples
- The expected counts can't be too small (<5). If they are smaller, observations should be grouped.
- The classes in the „continuous" version may be chosen arbitrarily, but it is best if the theoretical probabilities are balanced.

Chi-squared test of independence

Model: (X_1, Y_1), ..., (X_n, Y_n) are an IID sample from a two-dimensional distribution with r*s values (denoted by the set $\{1, ..., r\} \times \{1, ..., s\}$).

Let the theoretical distribution be

$$p_{ij} = P(X = i, Y = j) \quad i = 1, \ldots, r \quad j = 1, \ldots, s$$

Denote $$p_{i\bullet} = \sum_{j=1}^{s} p_{ij}, \quad p_{\bullet j} = \sum_{i=1}^{r} p_{ij}$$

We want to veryfy independence of X and Y:

H_0: $p_{ij} = p_{i\bullet} * p_{\bullet j} \quad i = 1, \ldots, s, \; j = 1, \ldots, r$

H_1: $\neg H_0$

Chi-squared test of independence – cont.

The empirical distribution may be summarized by a table (so-called contingency table, or crosstab)

$i \setminus j$	1	2	...	s	$N_{i\bullet}$
1	N_{11}	N_{12}		N_{1s}	$N_{1\bullet}$
2	N_{21}	N_{22}		N_{2s}	$N_{2\bullet}$
...					
r	N_{r1}	N_{r2}		N_{rs}	$N_{r\bullet}$
$N_{\bullet j}$	$N_{\bullet 1}$	$N_{\bullet 2}$		$N_{\bullet s}$	n

Chi-squared test of independence – cont. (2)

This is a special case of a goodness-of-fit test with $(r\text{-}1) + (s\text{-}1)$ parameters to be estimated:

The test statistic:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{ij} - N_{i\bullet} N_{\bullet j}/n)^2}{N_{i\bullet} N_{\bullet j}/n}$$

has a chi-squared distribution with $(r\text{-}1)(s\text{-}1)$ degrees of freedom (if H_0 is true)

Chi-squared test of independence – example

We verify independence of political and musical preferences, for signif. level α =0.05

	Support X	Do not support X	Total
Listen to jazz	25	10	35
Listen to rock	20	20	40
Listen to hip-hop	15	10	25
Total	60	40	100

$$\chi^2 = \frac{(25-60*35/100)^2}{60*35/100} + \frac{(20-60*40/100)^2}{60*40/100} + \frac{(15-60*25/100)^2}{60*25/100} + \frac{(10-40*35/100)^2}{40*35/100} + \frac{(20-40*40/100)^2}{40*40/100} + \frac{(10-40*25/100)^2}{40*25/100} \approx 3.57$$

$$\chi^2_{1-0.05}((2-1)(3-1)) = \chi^2_{0.95}(2) \approx 5.99$$

→ **no grounds to reject H_0.**

Warsaw University
Faculty of Economic Sciences

Source: W. Niemiro

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).
- 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 **a priori/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:

$$f(x_1, x_2, \ldots, x_n, \theta) = f(x_1, x_2, \ldots, x_n \mid \theta)\pi(\theta)$$

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

$$\pi(\theta \mid x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n \mid \theta)\pi(\theta)}{m(x_1, \ldots, x_n)},$$

where $m(x_1, \ldots, x_n) = \int_\Theta f(x_1, \ldots, x_n \mid \theta)\pi(\theta)d\theta$

is a marginal distribution for the obs.

Bayesian Model – a posteriori distribution

$\pi(\theta \mid x_1,\ldots,x_n)$ is called the **a posteriori/ 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

A priori and a posteriori distributions: examples

1. Let $X_1, ..., X_n$ 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 $\theta \in (0,1)$

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

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

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

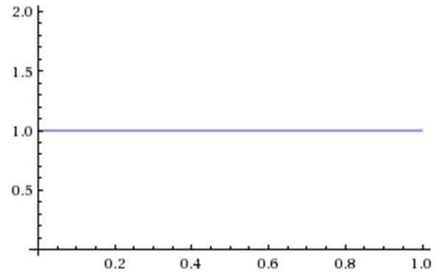
then the a posteriori distribution:

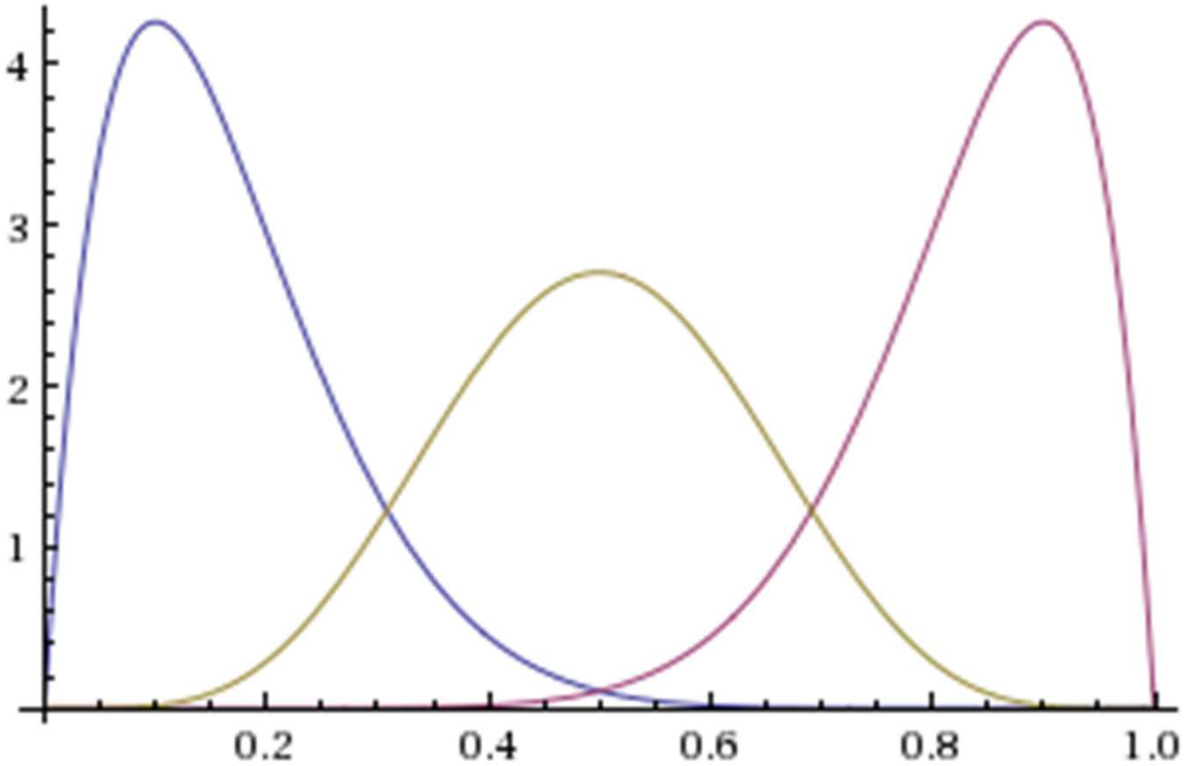
$$\text{Beta}\left(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta\right)$$

Warsaw University
Faculty of Economic Sciences

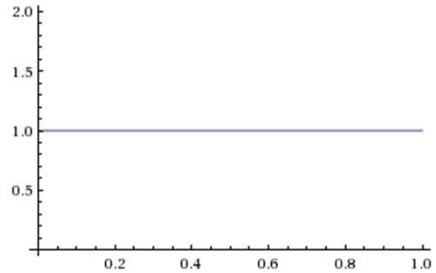
conjugate prior for Bernoulli distr.

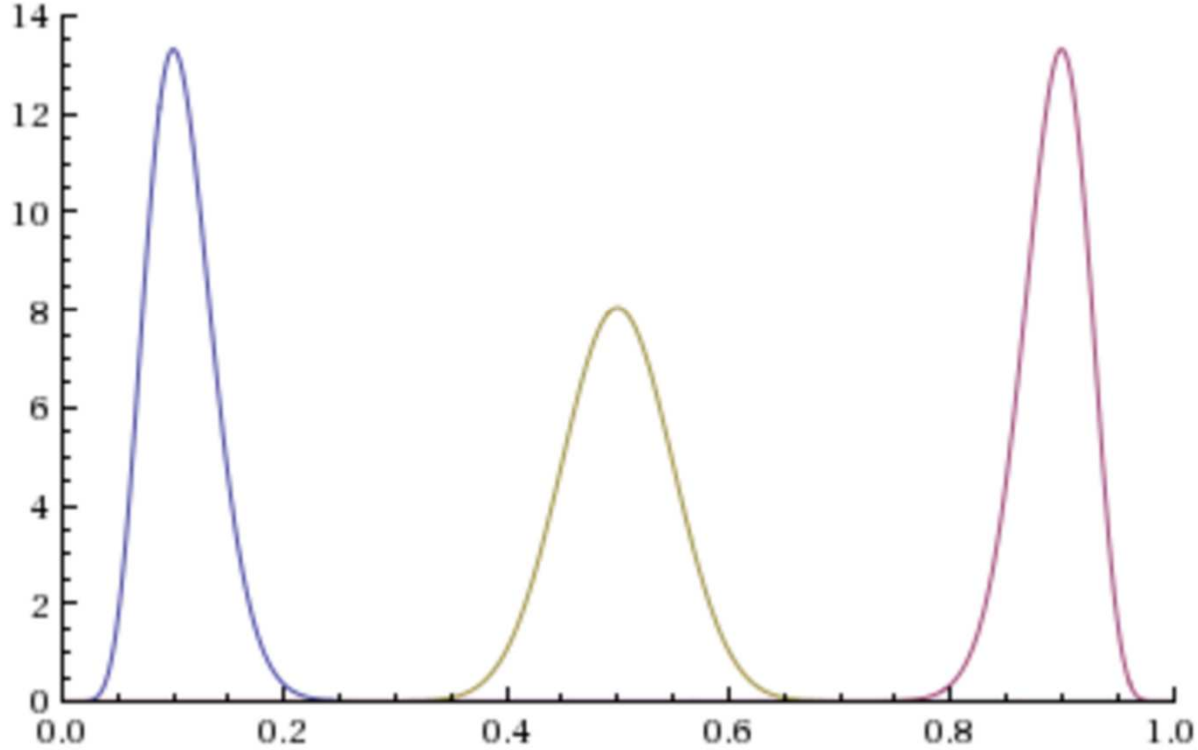
For a Beta (1,1) prior and data: n=10 and 1, 5, 9 successes



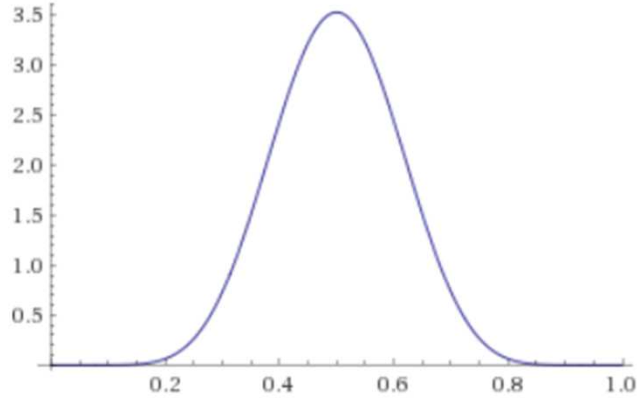


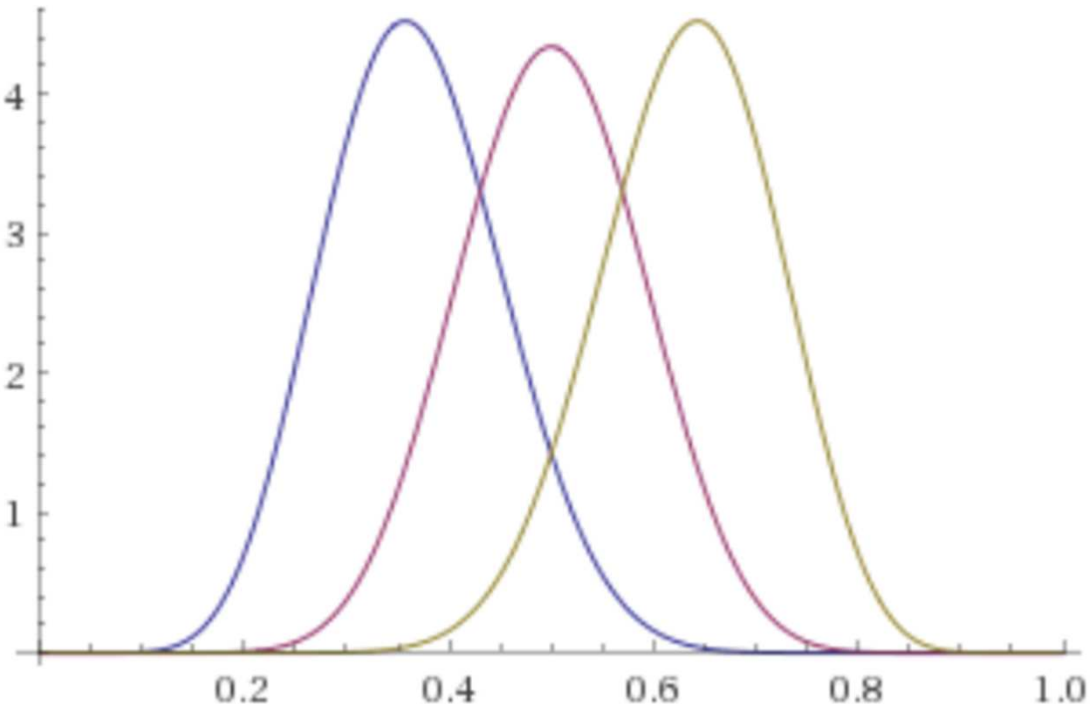
For a Beta (1,1) prior and data: n=100 and 10, 50, 90 successes



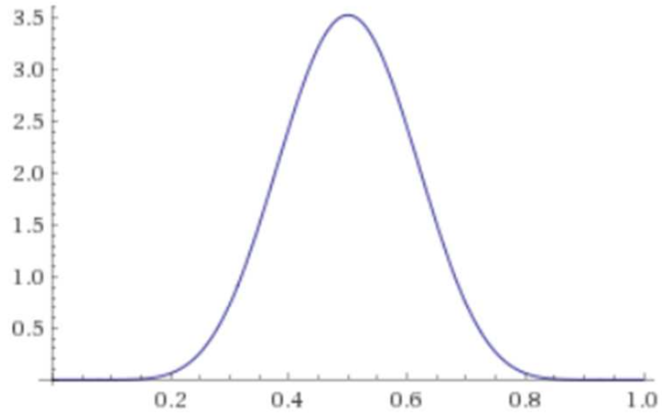


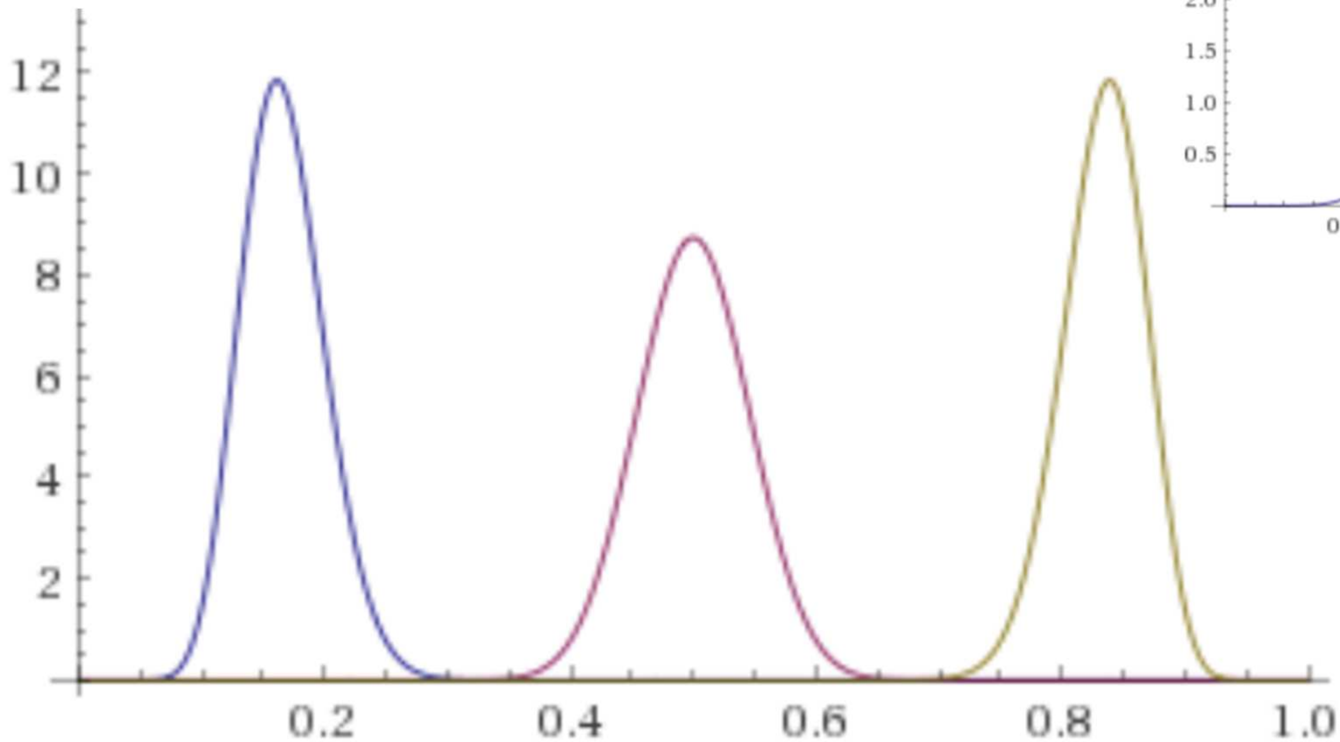
For a Beta (10,10) prior and data: n=10 and 1, 5, 9 successes



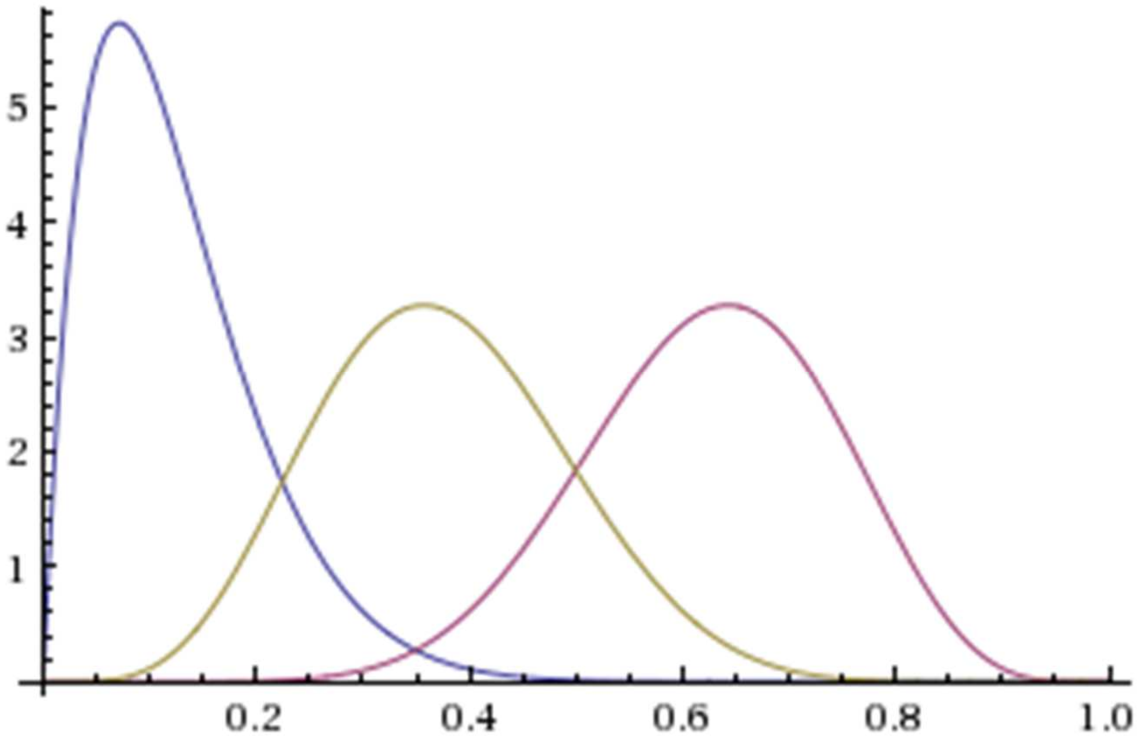


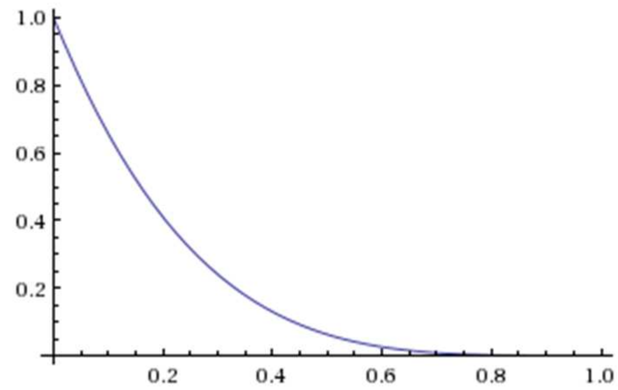
For a Beta (10,10) prior and data: n=100 and 10, 50, 90 successes



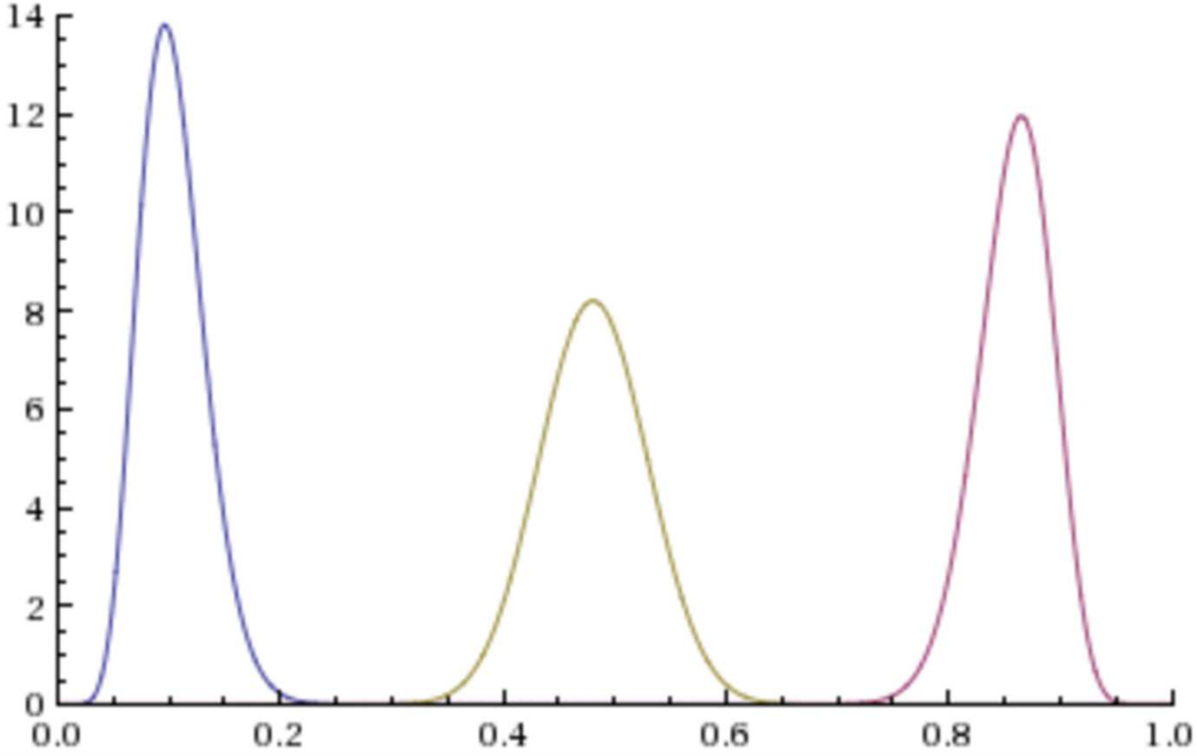


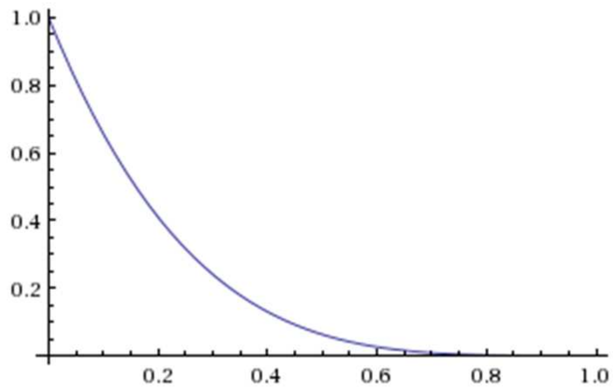
For a Beta (1,5) prior and data: n=10 and 1, 5, 9 successes





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





A priori and a posteriori distributions: examples (2)

2. Let $X_1, \ldots, X_n$ be IID r.v. from $N(\theta, \sigma^2)$, and σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the posterior distribution for θ:

$$N\left(\frac{n\frac{1}{\sigma^2}\overline{X} + \frac{1}{\tau^2}m}{n\frac{1}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \right)$$

conjugate prior for a normal distr.

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) / Maximum a posteriori Probability (MAP) estimate

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

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

i.e.

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

BMP: examples

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

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

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

We know the a posteriori distribution:

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

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 an a priori U(0,1) (i.e. Beta(1,1) distr.), we have $BMP(\theta)=5/10 = ½$

and for 9 successes in 10 trials for the same a priori distr., we have $BMP(\theta)=9/10$

BMP: examples (2)

2. Let $X_1, \ldots, X_n$ be IID r.v. from $N(\theta, \sigma^2)$, with σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the a posteriori distr for θ: $N\left(\frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$

so

$$BMP(\theta) = \frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}$$

i.e. if we have sa sample of 5 obs 1.2; 1.7 ; 1.9 ; 2.1; 3.1 from distr. N(θ, 4) and the a priori distr is $\theta \sim N(1, 1)$, then

$$BMP(\theta) = (5/4 * 2 + 1)/(5/4 + 1) = 14/9 \approx 1.56$$

and if the a priori distr were $\theta \sim N(3, 1)$, then

$$BMP(\theta) = (5/4 * 2 + 1*3)/(5/4 + 1) = 22/9 \approx 2.44$$

Bayes Estimator

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

$L(\theta, a)$ – **loss function,** depends on the true value of θ and the decision a.

e.g. if we want to estimate $g(\theta)$:

$L(\theta, a) = (g(\theta) - a)^2$ – quadratic loss function

$L(\theta, a) = |g(\theta) - a|$ – module loss function

Bayes Estimator – cont.

We can also define the **accuracy of an estimate** for a given loss function :

$$acc(\Pi,\hat{g}(x)) = E\left(L(\theta,\hat{g}(x)) \mid X = x\right) = \int_{\Theta} L(\theta,\hat{g}(x))\pi(\theta \mid x)d\theta$$

(the average loss of the estimator for a given a priori distribution and data, i.e. for a specific posterior distribution)

Bayes Estimator – cont. (2)

The **Bayes Estimator** for a given loss function $L(\theta, a)$ is $\hat{g}_B$ such that

$$\forall x \quad acc(\Pi, \hat{g}_B(x)) = \min_a acc(\Pi, a)$$

For a quadratic loss function $(\theta - \text{a})^2$:

$$\hat{\theta}_B = E(\theta \mid X = x) = E(\Pi_x)$$

For a module loss function $|\theta - \text{a}|^2$:

$$\hat{\theta}_B = Med(\Pi_x)$$

more generally: $E(g(\theta)|x)$

Bayes Estimator: Example (1)

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 poster distribution:

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

so the Bayes Estimator is

$$\hat{\theta}_B = \frac{\sum_{i=1}^{n} x_i + \alpha}{n + \beta + \alpha}$$

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

i.e. for 5 successes in 10 trials for an a priori U(0,1) (i.e. Beta(1,1) distr.), we have $\hat{\theta}_B$=6/12 = ½

and for 9 successes in 10 trials for the same a priori distr., we have $\hat{\theta}_B$=10/12 = 5/6

Warsaw University
Faculty of Economic Sciences

BMP: examples

1. Let $X_1, \ldots, X_n$ 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 poster distribution:

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

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

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 an a priori U(0,1) (i.e. Beta(1,1) distr.), we have $BMP(\theta)=5/10 = ½$

and for 9 successes in 10 trials for the same a priori distr., we have $BMP(\theta)=9/10$

Bayes Estimator: examples (2)

2. Let $X_1, ..., X_n$ be IID r.v. from $N(\theta, \sigma^2)$, with σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the a posteriori distr for θ: $N\left(\frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$

so

$$\hat{\theta}_B = \frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}$$

i.e. if we have sa sample of 5 obs 1.2; 1.7 ; 1.9 ; 2.1; 3.1 from distr. N(θ, 4) and the a priori distr is $\theta \sim N(1, 1)$, then

$\hat{\theta}_B$ = (5 /4 * 2 + 1)/(5/4 + 1) = 14/9 ≈ 1.56

and if the a priori distr were $\theta \sim N(3, 1)$, then

$\hat{\theta}_B$ = (5 /4 * 2 + 1*3)/(5/4 + 1) = 22/9 ≈ 2.44

BMP: examples (2)

2. Let $X_1, \ldots, X_n$ be IID r.v. from $N(\theta, \sigma^2)$, with σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the a posteriori distr for θ: $N\left(\frac{n\frac{1}{\sigma^2}\overline{X} + \frac{1}{\tau^2}m}{n\frac{1}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \right)$

so

$$BMP(\theta) = \frac{n\frac{1}{\sigma^2}\overline{X} + \frac{1}{\tau^2}m}{n\frac{1}{\sigma^2} + \frac{1}{\tau^2}}$$

i.e. if we have sa sample of 5 obs 1.2; 1.7 ; 1.9 ; 2.1; 3.1 from distr. N(θ, 4) and the a priori distr is $\theta \sim N(1, 1)$, then

$$BMP(\theta) = (5/4 * 2 + 1)/(5/4 + 1) = 14/9 \approx 1.56$$

and if the a priori distr were $\theta \sim N(3, 1)$, then

$$BMP(\theta) = (5/4 * 2 + 1*3)/(5/4 + 1) = 22/9 \approx 2.44$$