Topics for the exam in statistics: fall 2014/winter 2015 semester

Chance event, Probability

- definitions of probability by Laplace and Kolmogorov.
- probability of the union (sum) of events: $\mathcal{P}(A \bigcup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(AB)$
- conditional probability; definition of **independent events**

$$\mathcal{P}(A|B) = \mathcal{P}(A) \qquad \mathcal{P}(B|A) = \mathcal{P}(B)$$

$$\boxed{\mathcal{P}(AB) = \mathcal{P}(B)\mathcal{P}(A|B) = \mathcal{P}(B)\mathcal{P}(A)}$$

Random Variable (RV)

- Expected value of a RV: $E(X)$ – definitions for the case of a discrete and continuous RV.
- moments and basic parameters: **variance** (as the expected value of the square of deviations from ??); skewness (the measure of what property?)
- other parameters: quantile (sketch a graph showing the meaning of the – say – 95%-quantile: $q_{0.95}$); median, quartile.
- the probability function and the cummulative distribution function (the discrete RV); the probability density function $f(x)$ and the cummulative distribution function $F(x)$ (for the continuous RV). What are the meanings of these two functions in terms of probability? What are the relations linking $F(x)$ and $f(x)$ for the dicrete- and continuous-RV

Basic distributions of discrete and continuous RV,

their expected values $E(X)$, variances $\sigma^2(X)$:

- Poisson distribution ($E(X) = \sigma^2(X) = \lambda$. The basic formula for $\mathcal{P}; \lambda$). For ambitious (and well-trained in algebra) students: show by definition that $E(X) = \sum_{k=0}^{\infty} k \cdot \mathcal{P}(X = k; \lambda) = \lambda$ (you may find this derivation at my page)
- Bernoulli (binomial) distribution (definition, expected value, variance)
- Uniform distribution (definition, expected value, variance)
- Normal distribution (definition, expected value, variance)

Normal distribution

- the standardization procedure: if X follows a normal distribution with $E(X) = \mu$ and σ the standardized variable: $Z = \dfrac{X - \mu}{\sigma} \to N(0,1)$ – the standardized normal distribution.
- a rough graph of $N(0,1)$. What are the portions of the area under the graph for intervals 2-, 4-, and 6-σ wide?
- The Chebyshev theorem
- the Central Limit Theorem – what happens if we add together several (n) RVs that all follow the same type of distribution? What is the expected value of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$? And its variance – $\sigma^2(\bar{X})$?

Estimators

- what is 'statistical sample'?
- what is the commonly used estimator for $E(X)$?
- the same – for variance. Define S^2 and S^{*2} estimators. Why that (small) difference?

- if we have several values of RV x_i and their variances σ_i what estimator should be used for $E(X)$. Consult my page (at the very end of the presentation)

Confidence Intervals

- describe in detail procedures for constructing a – say – 95% confidence interval for the expected value. Assume: (a) a statistical sample of a rather large size ($\geqslant 30$); (b) a statistical sample with with $n \leqslant 10$. What are the formulae for the left- and right-hand boundary of those intervals (in both cases)?

Hypotheses testing

- what is 'rejection region' for the $H0$ hypothesis versus a given $H1$ one? Example: suppose you are to test the hypothesis $H0 : \mu = 5$ and your statistical sample has the $\bar{X} = 6$. σ is known and is equal to 3. Where will you put the rejection region if significance level $\alpha = 0,05$?
- describe the chi-squared test for testing a hypothesis about our RV conforming with the uniform distribution: we assume that in each of the unit intervals: $[0, 1), [1, 2), \ldots, [9, 10)$ we should have the same number of x-values $(n_i)_{theory}$ and our measurements give 10 values x-values $(n_i)_{experiment}$. How do we proceed? (revise the Mendel example) my page