Merkpunkte — 4.1.-6. Wahrscheinlichkeit I

1. Wahrscheinlichkeiten bilden ein Modell, das etwas über „unsichere” „Ereignisse” aussagt.
   Wahrscheinlichkeit = idealisierte relative Häufigkeit

2. Verknüpfung von Ereignissen mit Mengen-Operationen
   $\Rightarrow$ Regeln für entsprechende Wahrscheinlichkeiten

3. Zufallsvariable: Das Ergebnis eines „Versuchs” wird in Form einer oder mehrerer Zahlen festgehalten.
   Jargon: Jedem Elementarereignis ist (pro Zufallsvariable) ein Zahlenwert zugeordnet.

4. Verteilung einer diskreten Zufallsvariablen ist festgelegt, wenn man $P\langle X = x\rangle$ kennt für $x = 0, 1, 2, \ldots$.
   Gemeinsame Verteilung von 2 Zufallsvariablen:
   Alle $P\langle X = x, Y = y\rangle$ bekannt.

5. Zufallszahlen bilden Muster-Experiment zu einem beliebigen Modell.
   Wsch. $=$ rel. Häufigk. in „unendl." vielen Zufallszahlen.

6. Unabhängige Versuche, Ereignisse oder Zufallsvariable:
   - „beeinflussen sich gegenseitig nicht"
   - Wahrscheinlichkeiten multiplizieren sich:

   $$P\langle A \cap B\rangle = P\langle A\rangle \cdot P\langle B\rangle,$$
   $$P\langle X \in \widetilde{A},\ Y \in \widetilde{B}\rangle = P\langle X \in \widetilde{A}\rangle \cdot P\langle Y \in \widetilde{B}\rangle.$$

Merkpunkte

5.1-3. Diskrete Verteilungen I

1. Die Binomial-Verteilung eignet sich, wenn
   in n unabhängigen Versuchen jedesmal festgestellt wird,
   ob ein bestimmtes Ereignis („Erfolg") eintritt.
   $X \sim \mathcal{B}\langle n, \pi\rangle$.
   Param. π = W. für „Erfolg" in jedem Versuch.

2. Die Poisson-Verteilung eignet sich, wenn
   die Anzahl „Ereignisse" oder Punkte in einem bestimmten
   Zeit-Intervall, einer Fläche, einem Volumen gezählt wird.
   $X \sim \mathcal{P}\langle \lambda\rangle$.
   Parameter λ = „mittlere" Anz. = „Erwartungswert".

3. Merken Sie sich die möglichen Formen
   der Wahrscheinlichkeits-Stabdiagramme!

Merkpunkte Grundfragen der Schliessenden Statistik

1. Drei Grundfragen der Schliessenden Statistik:

   1 Welcher Parameter passt am besten zu den Beob.?

   $\longrightarrow$ Schätzung

   2 Sind Beob. mit best. Parameterwert θ_0 vereinbar?

   $\longrightarrow$ Test

   3 Welche Parameterwerte sind mit den Beob. vereinbar?

   $\longrightarrow$ Vertrauensintervall

2. Test: Antwort auf Frage (2) nach Schema:
   Man legt ein „extremes" Ereignis fest,
   das Wahrscheinlichkeit $\leq 5\%$ hat, falls $H_0 : \theta = \theta_0$ gilt.
   H_0 wird abgelehnt oder beibehalten.
   „Statistischer Widerspruchs-Beweis".

3. Vertrauendsintervall: Gehört zu einem bestimmten Test.
   Prinzip: Für jeden Parameterwert
   den entspr. Test durchführen.
   Binomial-V.: Diagramm in umgekehrter Richtung ablesen.

Merkpunkte — 4.5-9. Wahrscheinlichkeit II

1. Bedingte Wahrscheinlichkeit:

   - Teil-Informationen erlauben genauere Modelle $P\langle B|A\rangle = \frac{P\langle A\cap B\rangle}{P\langle A\rangle}$ resp. $P\langle Y = y \mid X = x\rangle = \ldots$,
   - werden durch Baumdiagramme anschaulich,
   - sind oft einfacher zu bestimmen und helfen dann, zu einem gesamten Modell zu kommen.

2. Satz von Bayes: aus $P\langle B|A\rangle$ und „apriori"-W. $P\langle A\rangle$ die „aposteriori"-W. $P\langle A|B\rangle$ ausrechnen.
   Unvollständiges Wissen verbessern.

Merkpunkte — Transformationen v. Daten

1. Lineare Transformation von Daten
   = Veränderung von Mess-Nullpunkt und -Einheit
   lässt die Form der Verteilung unverändert.

2. Oft geht man zur standardisierten Stichprobe über:
   Mittelwert 0 und Standardabweichung 1. Form bleibt.

3. Nichtlineare, monotone Transformationen,
   z.B. Logarithmus- und Wurzel-Transformation,
   verändern (verkleinern) die (positive) Schiefe.
   Können dazu führen, dass gleiche Differenzen
   als gleiche Unterschiede interpretiert werden können.

Merkpunkte **Stetige Verteilungen**

1. Verteilungen sind charakterisiert durch
   - kumulative Verteilungsfunktion $F^{(X)}\langle . \rangle$,
   - Dichte $f^{(X)}\langle . \rangle$: Ableitung von F
   (für diskrete Zufallsv: W. der einzelnen Werte p_x),
   - Erwartungswert $\mathcal{E}\langle X \rangle$ und Varianz $\text{var}\langle X \rangle$
   (und weitere Kennzahlen).

2. Gebräuchliche stetige Verteilungen sind:
   - Normalverteilung $\mathcal{N}\langle \mu, \sigma^2 \rangle$,
   - Lognormal-Verteilung,
   - Exponential-Verteilung $\mathcal{E}xp\langle \lambda \rangle$.

Merkpunkte

1. Zusammenhänge zw. 2 Variablen kann man beschreiben...
   - grafisch: Streudiagramm. (Funkt. nicht immer auf Anhieb!)
   - durch eine Korrelation,
   - durch Regression (wird später behandelt).
2. Die Pearson- (gewöhnliche) Korrelation charakterisiert den „linearen Anteil" des Zusammenhangs.
3. Die (Spearman'sche) Rang-Korrelation charakterisiert den „monotonen Anteil" des Zusammenhangs.
4. Eine hohe Korrelation reicht nicht, um auf einen Ursache–Wirkungs-Zusammenhang zu schliessen!

Merkpunkte **Gemeinsame Verteilungen**

1. Begriffe: Gemeinsame Vert., bedingte V., Unabhängigkeit

2. Wenn man die gemeinsame Verteilung kennt, kann man die Vert. der Summe zweier (mehrerer) Zufallsvariabler (oder einer allgemeineren Funktion) bestimmen.

3. Summiert man Zufallsvariable, so addieren sich auch
   - die Erwartungswerte,
   - die Varianzen, wenn die X_i unabhängig sind.

4. Modell der einfachen Zufalls-Stichprobe:
   $X_1, ..., X_n$, gleich verteilt und unabhängig

5. $\mathcal{E}\langle \overline{X} \rangle = \mu$ $\quad$ $\text{var}\langle \overline{X} \rangle = \sigma^2/n$

Merkpunkte

Wenn man zählt, wieviele Personen, Objekte, Beob.-Einheiten zu jeder von m Kategorien gehören, ist die Multinomiale Verteilung das geeignete Modell.
Dabei muss, wie bei der Binomial-Vert., vorausgesetzt werden, dass die Beobachtungseinheiten unabhängige Ergebnisse liefern.

Merkpunkte — **Summen und Funkt. v. Zufallsvar.**

1. Summiert man Zufallsvariable, so addieren sich auch
   - die Erwartungswerte,
   - die Varianzen, wenn die X_i unabhängig sind.
2. Sind die Summanden normal- oder Poisson oder binomial-verteilt mit gleichem π,
   so haben die Summen wieder diese Verteilungen
   (mit anderen Parametern).
3. Mit Simulation kann man für beliebige Funktionen
   von mehreren Zufallsvariablen die Verteilung bestimmen.
   (Oft geht es auch mit Mathematik, mind. näherungsweise.)

Merkpunkte **Tests II**

1. Grundfrage: Ist ein bestimmter Parameterwert
mit den Daten „verträglich”?

2. Rezept: Nullhypothese H_0, Alternative H_A
$\longrightarrow$ Teststatistik U = Zufallsvar., die
die Beob. zusammenfasst (oft Schätzung des Parameters)
Vert. von U unter H_0 (oft über Standardisierung $U \mapsto T$)
Kritischer Bereich mit W. $\alpha = 0.05$ (unter H_0.
Durchführung: Wert t der Teststatistik für die Daten
$\longrightarrow$ Entscheidung: H_0 abgelehnt oder beibehalten
(nie bewiesen!)
Alternativ: P-Wert aus t berechnen
$\longrightarrow$ Ablehnung, falls $p < 0.05$.

3. Tests für eine / 2 verbundene Stichprobe/n: z- und t-Test
   Besser: Vorzeichen-Rangsummen-Test von Wilcoxon

4. Tests für zwei unabhängige Stichproben: z- und t-Test
   Besser: Rangsummen-Test von Wilcoxon, Mann, Whitney

5. Tests sind nicht der Stein des Weisen.

Merkpunkte **Vertrauensintervalle**

1. Grundfrage: Welche Parameterwerte sind
   mit den Beobachtungen verträglich?
   $\longrightarrow$ Vertrauensintervall

2. Jede Test-Methode für einen Parameterwert $(H_0 : \theta = \theta_0)$ führt zu einer Vertrauensintervall-Methode

3. Es lebe der Unterschied zwischen einem Vertrauensintervall und einem Annahmebereich!

4. Vertrauensintervalle enthalten die nützlichste Information über einen Parameter!

Merkpunkte

1. Aus Zufallsvariablen werden oft neue berechnet. Die Genauigkeit des Resultats lässt sich via Linearisierung und die bekannten Formeln für lineare Funktionen und Summen von unabh. Var. näherungsweise recht einfach berechnen.

2. Nach dem Zentralen Grenzwertsatz sind Summen und Mittelwerte über viele Zufallsvariable näherungsweise normalverteilt mit (Erwartungswert und Varianz sind klar.)

3. („anständige”) Funktionen von vielen Zv. sind ebenfalls näherungsweise normalverteilt. Die Varianz ist schwieriger zu bestimmen.