Vorlesung 7 (Montag 26.2.2018)

6.4 Zurückweisungsmethode

Für (analytisch) nicht-integrable W.-Dichten, oder (analytisch) nicht-invertierbare Verteilungen.

Bedingung: W-Dichte $p(x)$ passt in Kasten $[x_0, x_1) \times [0, p_{\max}]$, d.h. $p(x) = 0$ für $x \notin [x_0, x_1]$ und $p(x) \leq p_{\max}$.

Grundidee: Erzeuge zufällige Paare (x, y), gleichverteilt in $[x_0, x_1) \times [0, p_{\max})$. Akzeptiere nur die x mit $y \leq p(x)$, d.h. die Paare unterhalb $p(x)$, siehe Abb. 16.

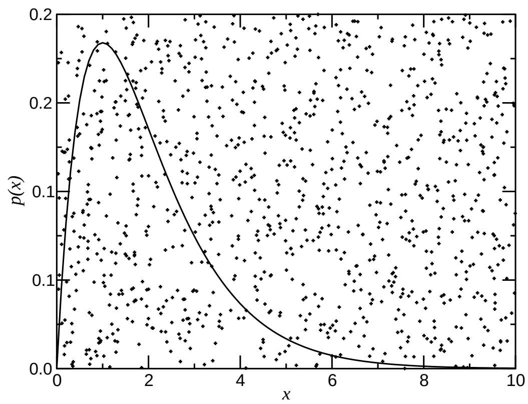

Figure 16: Zurückweisungsmethode: Punkte (x, y) sind gleichmäßig in dem Rechteck verteilt. Die Wahrscheinlichkeit, dass $y \leq p(x)$ ist proportional zu $p(x)$.

Implementierung als Funktion (Programm `reject.c`):

```
/** gerenates random number for 'pdf' in the range **/
/** ['x0', 'x1'). condition: pdf(x)<='p_max' in    **/
//* the range ['x0', 'x1')                          **/
double reject(double p_max, double x0, double x1,
              double (* pdf)(double))
{
  int found;                 /* flag if valid number has been found */
  double x,y;                 /* random points in [x0,x1]x[0,p_max] */
  found = 0;
  while(!found)                    /* loop until number is generated */
  {
    x = x0 + (x1-x0)*drand48();             /* uniformly on [x0,x1] */
    y = p_max *drand48();                 /* uniformly in [0,p_max] */
    if(y <= pdf(x))                                     /* accept ? */
        found = 1;
  }
  return(x);
}
```

Beispiel:

```
/** artifical pdf **/
double pdf(double x)
{
  if( (x<0)||
      ((x>=0.5)&&(x<1))||
      (x>1.5) )
      return(0);
  else if((x>=0)&&(x<0.5))
      return(1);
  else
      return(4*(x-1));
}
```

ergibt bei 100000 Zufallszahlen:
Nachteil: Es müssen unter Umständen viele Zufallszahlen weggeworfen werden. Effizienz $1/(2p_{\max}(x_1-x_0))$. (Faktor 1/2 da man mind. 2 Zahlen (x,y) für eine Zufallszahl braucht).

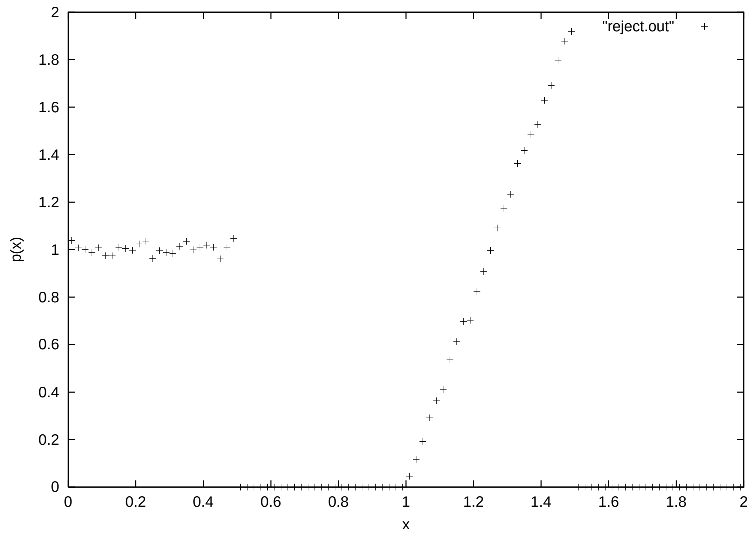

Figure 17: Zurückweisungsmethode: Histogramm für die angegebene W. Dichte.

6.5 Gaußverteilung

Gaußverteilung mit Erwatungswert m und Standardabweichung σ, die am meistverbreitete Verteilung in Physiksimulationen. W.-Dichte (siehe auch Abb. 18):

$$p_G(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z-m)^2}{2\sigma^2}\right) \tag{37}$$

Im Folgenden: z sei standard-normalverteilt ($m = 0,\ \sigma = 1$). Allgemeiner Fall: Benutze $\sigma z + m$.

[Selbsttest]

Nennen Sie Möglichkeiten (nährungsweise) normalverteilte Werte zu erzeugen

---

Hier:
<u>Box-Müller Methode</u>: Nehme zwei $[0,1)$ gleichverteilte Zahlen u_1, u_2 und setzte:

$$\begin{aligned} n_1 &= \sqrt{-2\log(1-u_1)}\cos(2\pi u_2) \\ n_2 &= \sqrt{-2\log(1-u_1)}\sin(2\pi u_2) \end{aligned}$$

dann sind n_1, n_2 standard-normalverteilt.
Beweis [4, 2]: Wir schreiben n_1, n_2 in Polarkoordinaten (r, θ), d.h. $(r, \theta) = f(n_1, n_2)$, das Inverse ist:

$$\begin{aligned} n_1 &= r\cos(\theta) \\ n_2 &= r\sin(\theta) \end{aligned}$$

Gesucht: W.-Dichte für (r, θ).

Allgemeiner Fall $(W, Z) = f(X, Y)$ (also $(X, Y) \leftrightarrow (n_1, n_2)$), $p_{X,Y}$ sei die (gemeinsame) W. Dichte für (X, Y) . Dann gilt:

$$p_{W,Z}(w, z) = p_{X,Y}(f^{-1}(w, z))|\mathbf{J^{-1}}|\,, \tag{38}$$

mit $|\mathbf{J^{-1}}|$ ist die Jacobi Determinante der inversen Transformation.

Hier ist:

$$|\mathbf{J^{-1}}| = \left| \begin{array}{cc} \frac{\partial n_1}{\partial r} & \frac{\partial n_1}{\partial \theta} \\ \frac{\partial n_2}{\partial r} & \frac{\partial n_2}{\partial \theta} \end{array} \right| = \left| \begin{array}{cc} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{array} \right| = r\cos^2(\theta) + r\sin^2(\theta) = r \tag{39}$$

Da n_1, n_2 Gauß-verteilt sein sollen:

$$p_{R,\Theta}(r, \theta) = \frac{r}{2\pi} e^{-n_1^2/2 - n_2^2/2} = \frac{r}{2\pi} e^{-r^2/2} \tag{40}$$

Faktorisiert! Wir nehmen θ gleichförmig in $[0, 2\pi)$ verteilt (d.h. Erzeugung $\theta = 2\pi u_2$). Es bleibt $p_R(r) = re^{-r^2/2}$ (*) übrig. Wie erzeugt man Zufallszahlen gemäß p_R?

Verteilungsfunktion (leicht zu integrieren) $F_R(r) = 1 - \exp(-r^2/2)$ $(r \geq 0)$ $\rightarrow$ Inversion $r = \sqrt{-2\log(1-u)}$ $\rightarrow$ gewünschte Verteilung für r. QED.

6.6 Grundlagen Datenanalyse

Gegeben: n Messpunkte ("Stichprobe") $\{x_0, x_1, \ldots, x_{n-1}\}$

Problem: zugrundeliegende Verteilung $F(x)$ meistens unbekannt.

6.6.1 Schätzwerte

Schätzwerte $h = h(x_0, x_1, \ldots, x_{n-1})$ sind selber Zufallsgrößen: $H = h(X_0, X_1, \ldots, X_{n-1})$

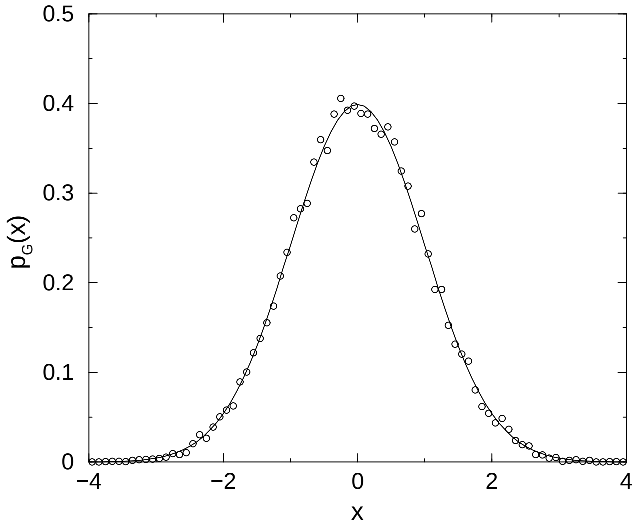

Figure 18: Gaußverteilung mit Mittelwert 0 und Varianz 1. Die Kreise entsprechen einem Histogramm, das aus 10^4 Box-Müller erzeugten Zufallszahlen erstellt wurde.

- Mittelwert (MW)

$$\overline{x} \equiv \frac{1}{n} \sum_{i=0}^{n-1} x_i \tag{41}$$

- Stichprobenvarianz

$$s^2 \equiv \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \overline{x})^2 \tag{42}$$

- Stichproben Standardabweichung

$$s \equiv \sqrt{s^2} \tag{43}$$

MW: Zur Schätzung des Erwartungswertes $\mu = \mathrm{E}[X]$. MW entspricht ZV $\overline{X} = \frac{1}{n}\sum_{i=0}^{n-1} X_i$. $\Rightarrow$

$$\mu_{\overline{X}} \equiv \mathrm{E}[\overline{X}] = \mathrm{E}\left[\frac{1}{n}\sum_{i=0}^{n-1} X_i\right] = \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}[X_i] = \frac{1}{n} n\, \mathrm{E}[X] = \mathrm{E}[X] = \mu \tag{44}$$

$\rightarrow$ der Mittelwert ist erwartungstreu.
Verteilung der $\overline{X}$ hat Varianz:

$$\begin{aligned}\sigma^2_{\overline{X}} &\equiv \mathrm{Var}[\overline{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=0}^{n-1} X_i\right] \overset{\mathrm{Var}[\alpha X]=\alpha^2\,\mathrm{Var}[X]}{=} \frac{1}{n^2}\sum_{i=0}^{n-1}\mathrm{Var}[X_i] \\ &= \frac{1}{n^2} n\,\mathrm{Var}[X] = \frac{\sigma^2}{n} \end{aligned} \tag{45}$$

$\rightarrow$ wird schmaler für wachsendes n
$\rightarrow$ Schätzung wird genauer (wobei σ^2 unbekannt)

$\rightarrow$ gesucht: erwartungstreuer Schätzwert für σ^2 Versuch für $S^2 = \frac{1}{n}\sum_{i=0}^{n-1}(X_i - \overline{X})^2$:

$$\begin{aligned}\mathrm{E}[S^2] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=0}^{n-1}(X_i-\overline{X})^2\right] = \mathrm{E}\left[\frac{1}{n}\sum_{i=0}^{n-1}(X_i^2 - 2X_i\overline{X} + \overline{X}^2)\right] \\ &\overset{\sum_i X_i = n\overline{X}}{=} \frac{1}{n}\left(\sum_{i=0}^{n-1}\mathrm{E}[X_i^2] - n\,\mathrm{E}[\overline{X}^2]\right) \overset{\mathrm{E}[Y^2]=\sigma_Y^2+\mu_Y^2}{=} \frac{1}{n}\big(n(\sigma^2+\mu^2) - n(\sigma^2_{\overline{X}} + \mu^2_{\overline{X}})\big) \\ &\overset{\sigma^2_{\overline{X}}=\frac{\sigma^2}{n}}{=} \frac{1}{n}\left(n\sigma^2 + n\mu^2 - n\frac{\sigma^2}{n} - n\mu^2\right) = \frac{n-1}{n}\sigma^2 \end{aligned} \tag{46}$$

S^2 ist <u>nicht</u> erwartungstreu, wohl aber $\frac{n}{n-1}S^2$

6.6.2 Fehlerbalken

Definition: <u>Mittlerer quadratischer Fehler</u> (MQF, engl. MSE) eines Schätzwertes $H = h(X_0, X_1, \dots, X_{n-1})$ für den H zugeordneten Parameter θ

$$\begin{aligned}\mathrm{MSE}(H) &\equiv \mathrm{E}[(H-\theta)^2] = \mathrm{E}[(H - \mathrm{E}[H] + \mathrm{E}[H] - \theta)^2] \\ &= \mathrm{E}[(H-\mathrm{E}[H])^2] + \mathrm{E}[2(H-\mathrm{E}[H])(\mathrm{E}[H]-\theta)] + \mathrm{E}[(\mathrm{E}[H]-\theta)^2] \\ &= \mathrm{E}[(H-\mathrm{E}[H])^2] + 2\underbrace{(\mathrm{E}[H]-\mathrm{E}[H])}_{=0}(\mathrm{E}[H]-\theta) + (\mathrm{E}[H]-\theta)^2 \\ &= \mathrm{Var}[H] + (\mathrm{E}[H]-\theta)^2 \end{aligned} \tag{47}$$

Falls Schätzwert erwartungstreu ($\mathrm{E}[H] = \theta$), Fehler durch die Varianz gegeben, nutze ($\sigma^2_{\overline{X}} = \sigma^2/n$) und (46).

Nachteile: Varianz a priori unbekannt (aber: siehe Resampling); MQF hat keine probabilistische Interpretation.

$\rightarrow$ Für Mittelwert (ohne Beweis):

$$\boxed{P\left(\overline{X} - z\frac{S}{\sqrt{n-1}} \le \mu \le \overline{X} + z\frac{S}{\sqrt{n-1}}\right) \approx 1 - \alpha} \qquad (48)$$

wobei $\alpha = \alpha(z)$ die Signifikanz, mit $\alpha = 0.32, 0.05, 0.003$ für $z = 1, 2, 3$.

Forgeschrittene Themen (siehe [5])

- Resampling (kommt später)
- Hypothesentest, insbesondere Vergleich von Verteilungen (χ^2 Test, Kolmogorov-Smirnov Test)
- Korrelationen
- Maximum-likelihood Methoden
- Daten fitten (siehe Gnuplot)