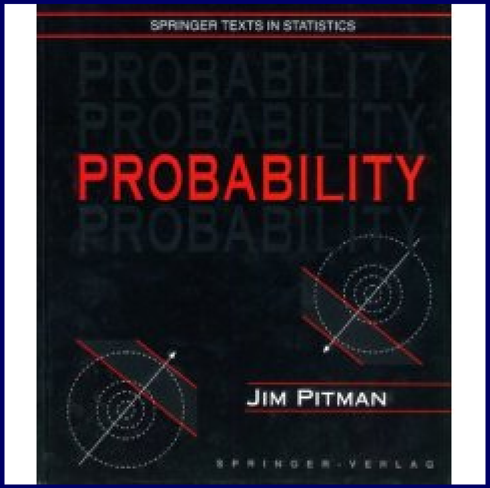
Introduction to probability

Stat 134 FAll 2005

Berkeley



Lectures prepared by:
Elchanan Mossel
elena Shvets

Follows Jim Pitman's book:

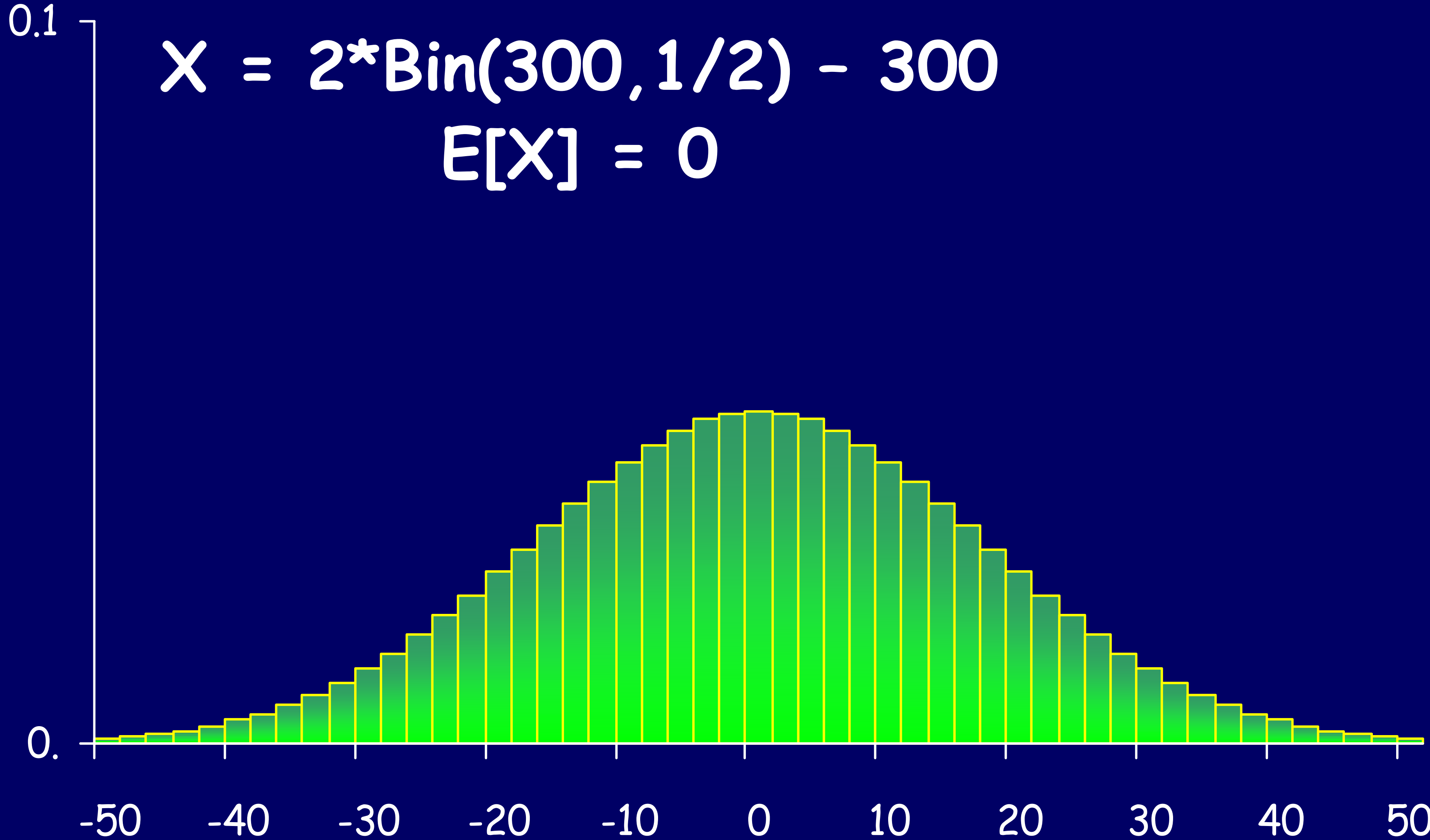
Probability

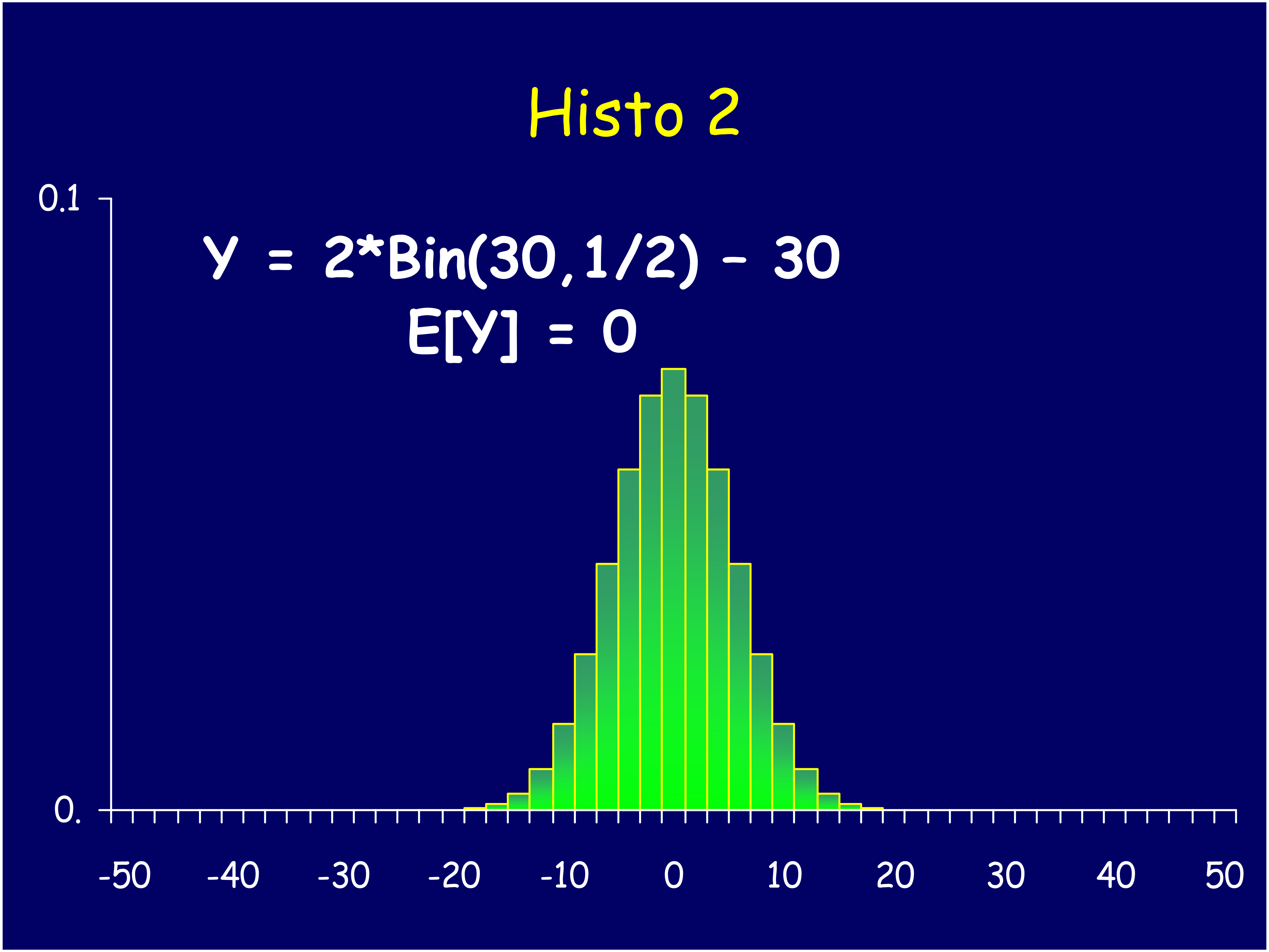
Section 3.3

Histo 1

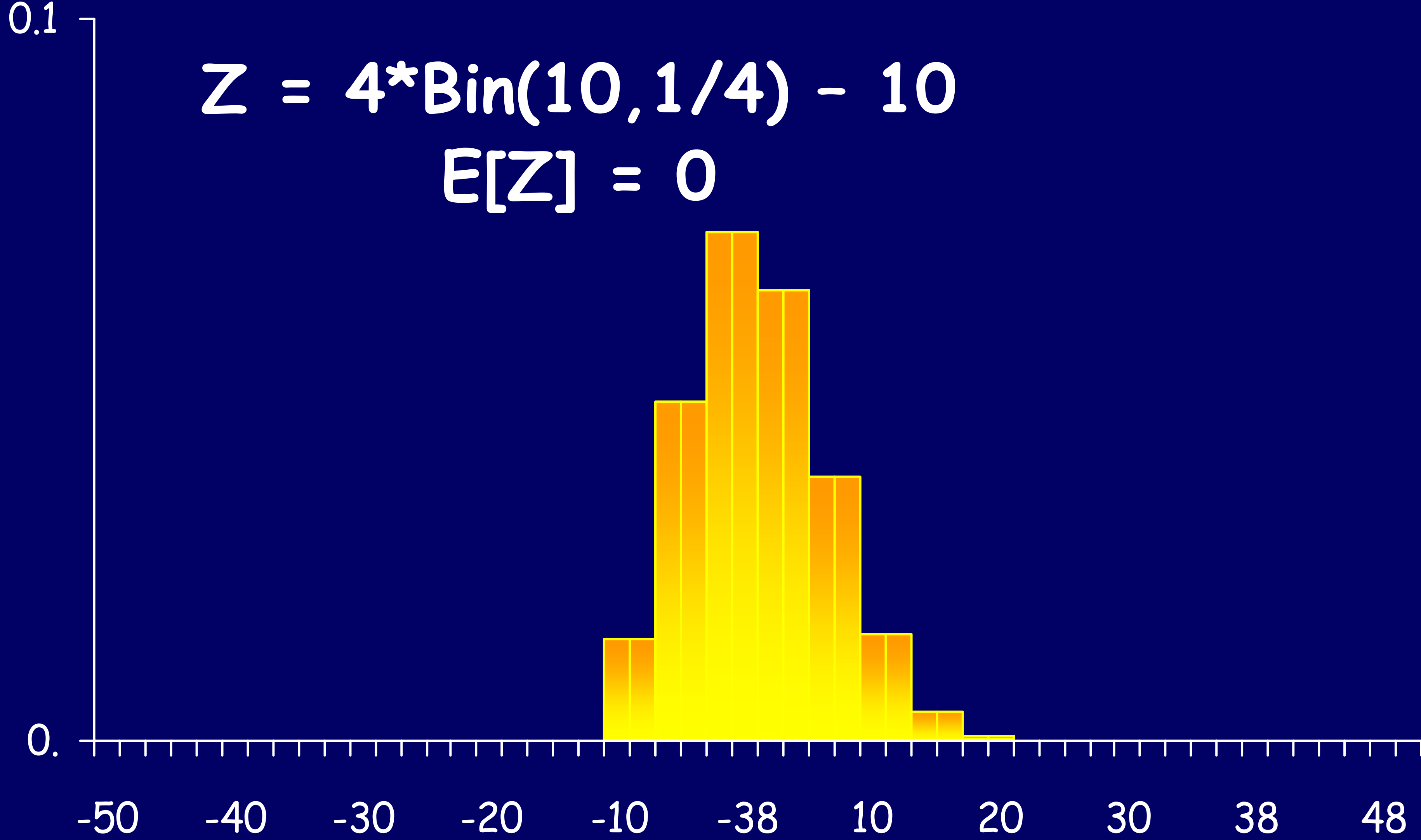
$X = 2*Bin(300, 1/2) - 300$

$E[X] = 0$

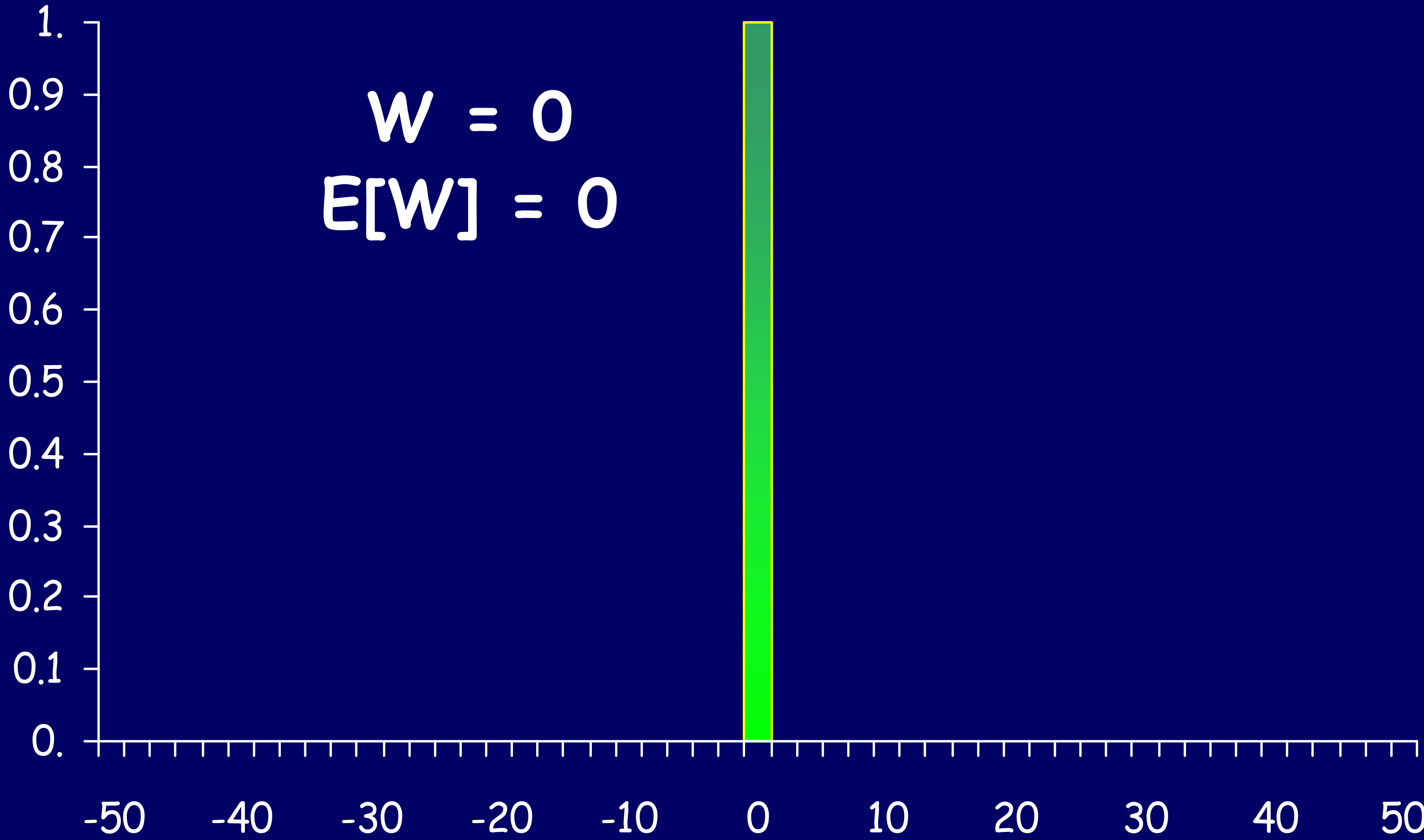



Histo 2
0.1
Y = 2*Bin(30,1/2) - 30
E[Y] = 0
0.
-50
-40
-30
-20
-10
0
10
20
30
40
50

Histo 3



Histo 4



A natural question:

- Is there a good parameter that allow to distinguish between these distributions?
- Is there a way to measure the spread?

Variance and Standard Deviation

- The variance of X, denoted by Var(X) is the mean squared deviation of X from its expected value $\mu = E(X)$:

$$Var(X) = E[(X-\mu)^2].$$

The standard deviation of X, denoted by SD(X) is the square root of the variance of X:

$$SD(X) = \sqrt{Var(X)}.$$

Computational Formula for Variance

Claim:

$$Var(X) = E(X^2) - E(X)^2.$$

Proof:

$$E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2]$$

$$= E[X^2] - 2\mu E[X] + \mu^2$$

$$= E[X^2] - 2\mu^2 + \mu^2$$

$$= E[X^2] - E[X]^2$$

Variance and SD

For a general distribution Chebyshev inequality states that for every random variable X, X is expected to be close to E(X) give or take a few SD(X).

Chebyshev Inequality:

For every random variable X and for all $k > 0$:

$$P(|X - E(X)| \geq k\, SD(X)) \leq 1/k^2.$$

Properties of Variance and SD

1. <u>Claim:</u> Var(X) ¸ 0.

   <u>Pf:</u> $Var(X) = \sum (x-\mu)^2 P(X=x)$ ¸ 0

2. <u>Claim:</u> $Var(X) = 0$ iff $P[X=\mu] = 1$.

Chebyshev's Inequality

$$P(|X - E(X)| \geq k\, SD(X)) \leq 1/k^2$$

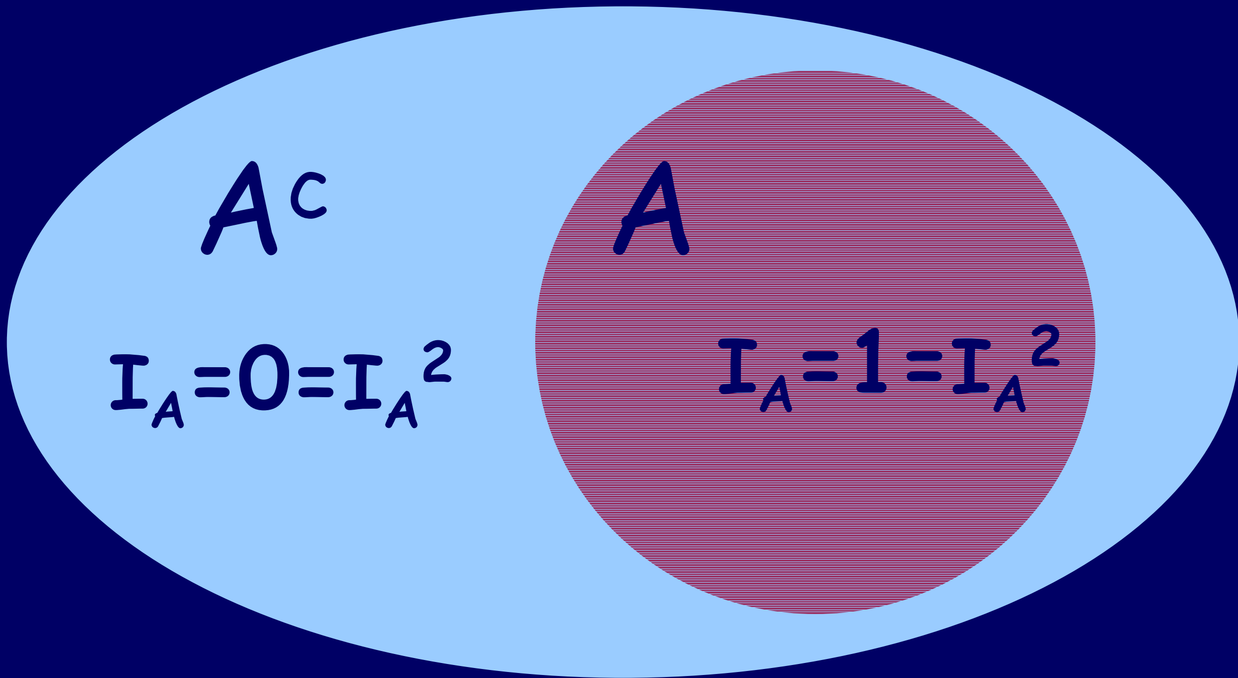
proof:

- Let $\mu = E(X)$ and $\sigma = SD(X)$.
- Observe that $|X-\mu| \geq k\sigma$, $|X-\mu|^2 \geq k^2 \sigma^2$.
- The RV $|X-\mu|^2$ is non-negative, so we can use Markov's inequality:
- $P(|X-\mu|^2 \geq k^2 \sigma^2) \leq E[|X-\mu|^2] / k^2 \sigma^2$

  $\leq \sigma^2 / k^2 \sigma^2 = 1/k^2$.

Variance of Indicators

Suppose I_A is an indicator of an event A with probability p. Observe that $I_A^2 = I_A$.



$$E(I_A^2) = E(I_A) = P(A) = p, \text{ so:}$$

$$Var(I_A) = E(I_A^2) - E(I_A)^2 = p - p^2 = p(1-p).$$

Variance of a Sum of Independent Random Variables

Claim: if $X_1, X_2, \ldots, X_n$ are independent then:

$$Var(X_1+X_2+\ldots+X_n) = Var(X_1)+Var(X_2)+\ldots+Var(X_n).$$

Pf: Suffices to prove for 2 random variables.

$E[(X+Y - E(X+Y))^2] = E[(X-E(X) + Y-E(Y))^2] =$

$E[(X-E(X))^2 + 2\, E[(X-E(X))\,(Y-E(Y))] + E(Y-E(Y))^2]=$

$Var(X) + Var(Y) + 2\, E[(X-E(X))]\; E[(Y-E(Y))]$ (mult.rule) $=$

$Var(X) + Var(Y) + 0$

Variance and Mean under scaling and shifts

- **Claim:** **$SD(aX + b) = |a|\ SD(X)$**
- Proof:

$Var[aX+b] = E[(aX+b - a\mu - b)^2] =$

$= E[a^2(X-\mu)^2] = a^2\sigma^2$

- Corollary: If a random variable X has
- $E(X) = \mu$ and $SD(X) = \sigma > 0$, then
- $X^* = (X-\mu)/\sigma$ has
- $E(X^*) = 0$ and $SD(X^*) = 1$.

Square Root Law

Let $X_1, X_2, \ldots, X_n$ be independent random variables with the same distribution as X, and let S_n be their sum:

$S_n = \sum_{i=1}^{n} X_i$, and

$$\overline{X} = \frac{S_n}{n}$$

their average, then:

$$E(S_n) = nE(X)$$

$$Var(S_n) = nVar(X)$$

$$SD(S_n) = \sqrt{n}SD(X).$$

$$E(\bar{X}_n) = E(X)$$

$$SD(\bar{X}_n) = \frac{SD(X)}{\sqrt{n}}$$

Weak Law of large numbers

<u>Thm:</u> Let $X_1, X_2, \ldots$ be a sequence of independent random variables with the same distribution. Let μ denote the common expected value $\mu = E(X_i)$. And let $\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$.

Then for every $\varepsilon > 0$:

$$P(|\overline{X}_n - \mu| < \varepsilon) \to 1 \text{ as } n \to \infty.$$

Weak Law of large numbers

Proof: Let $\mu = E(X_i)$ and $\sigma = SD(X_i)$. Then from the square root law we have:

$$E(\overline{X}_n) = \mu \text{ and } SD(\overline{X}_n) = \frac{\sigma}{\sqrt{n}}.$$

Now Chebyshev inequality gives us:

$$P(|\overline{X}_n - \mu| \geq \varepsilon) = P(|\overline{X}_n - \mu| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\frac{\sigma}{\sqrt{n}}) \leq \left(\frac{\sigma}{\varepsilon\sqrt{n}}\right)^2$$

For a fixed ε right hand side tends to 0 as n tends to 1.

The Normal Approximation

- Let $S_n = X_1 + \ldots + X_n$ be the sum of independent random variables with the same distribution.
- Then for large n, the distribution of S_n is approximately normal with mean $E(S_n) = n\,\mu$ and $SD(S_n) = \sigma\, n^{1/2}$,
- where $\mu = E(X_i)$ and $\sigma = SD(X_i)$.

In other words:

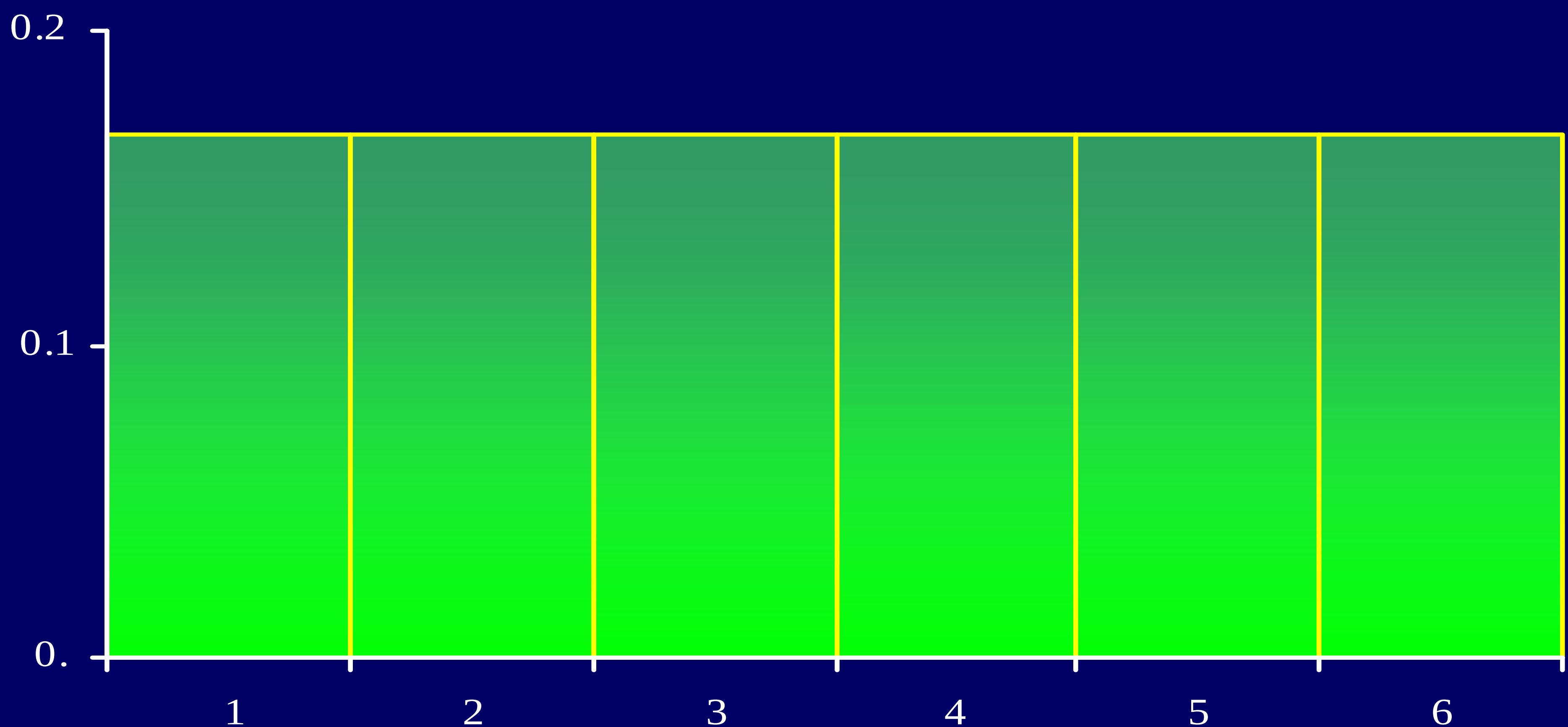
$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \sim \Phi(b) - \Phi(a)$$

Sums of iid random variables

Suppose X_i represents the number obtained on the i'th roll of a die.

Then X_i has a uniform distribution on the set $\{1,2,3,4,5,6\}$.

Distribution of X_1



Sum of two dice

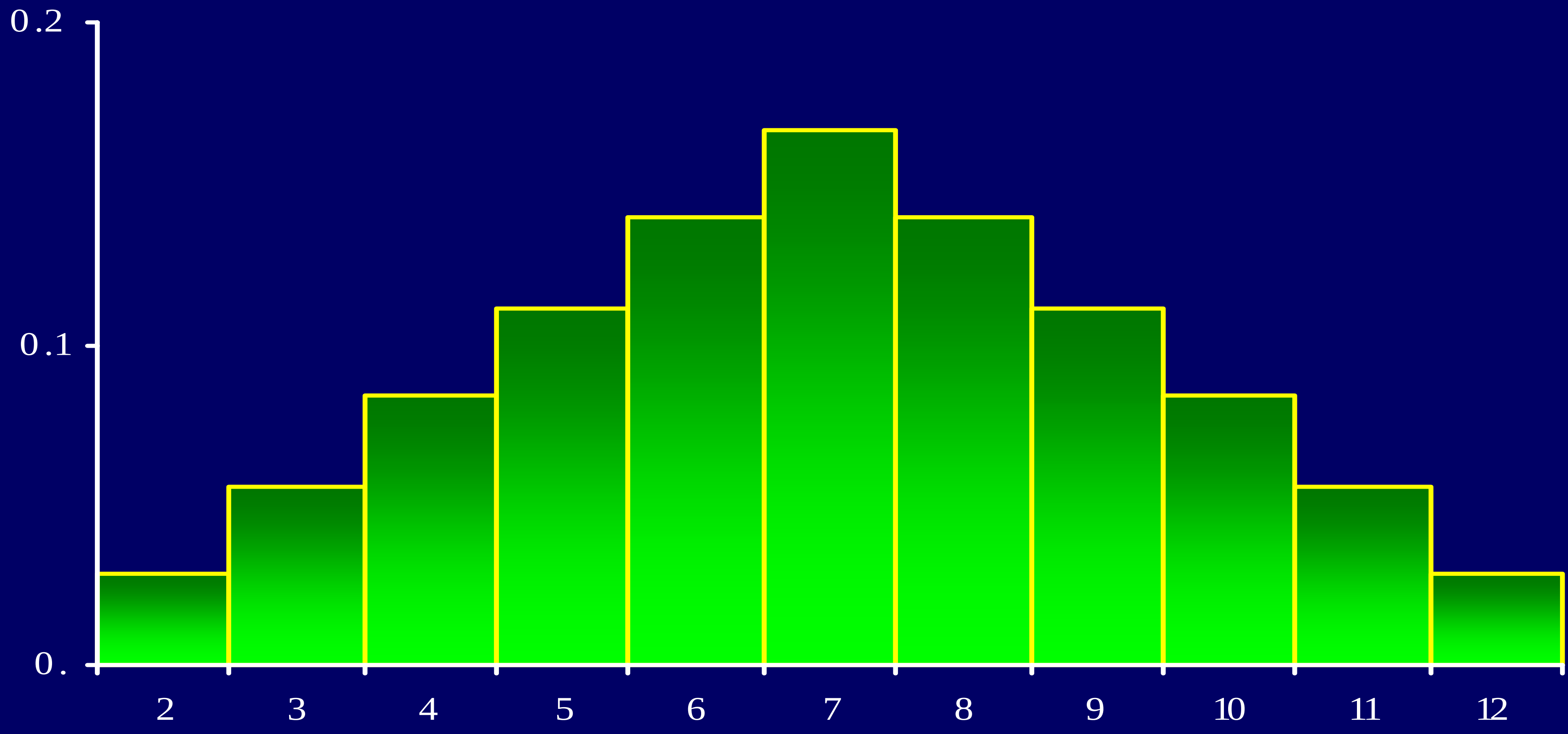
We can obtain the distribution of $S_2 = X_1 + X_2$ by the convolution formula:

$$P(S_2 = k) = \sum_{i=1}^{k-1} P(X_1=i)\, P(X_2=k-i \mid X_1=i),$$

by independence

$$= \sum_{i=1}^{k-1} P(X_1=i)\, P(X_2=k-i).$$

Distribution of S_2



Sum of four dice

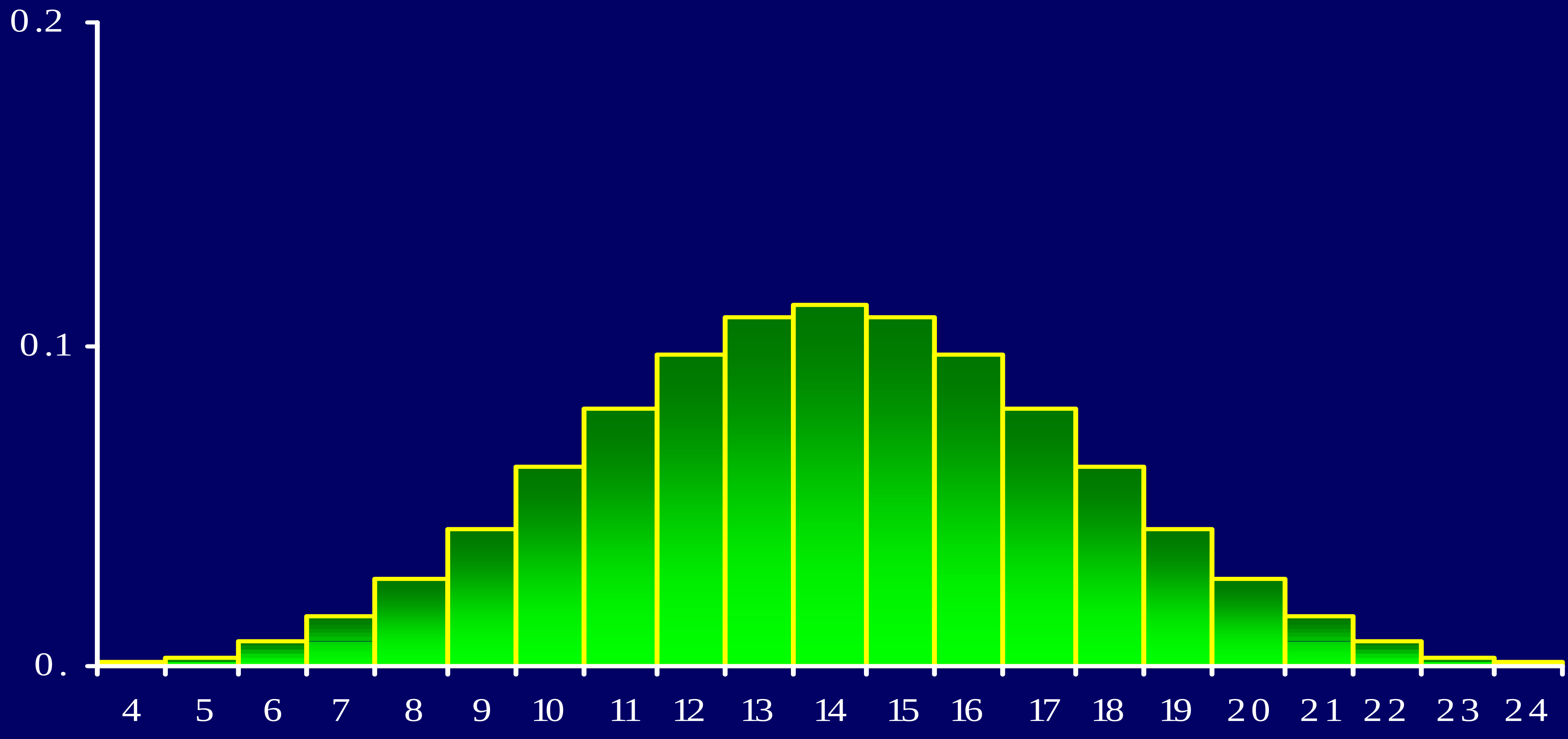
We can obtain the distribution of $S_4 = X_1 + X_2 + X_3 + X_4 = S_2 + S'_2$ again by the convolution formula:

$$P(S_4 = k) = \sum_{i=1}^{k-1} P(S_2=i)\, P(S'_2=k-i \mid S_2=i),$$

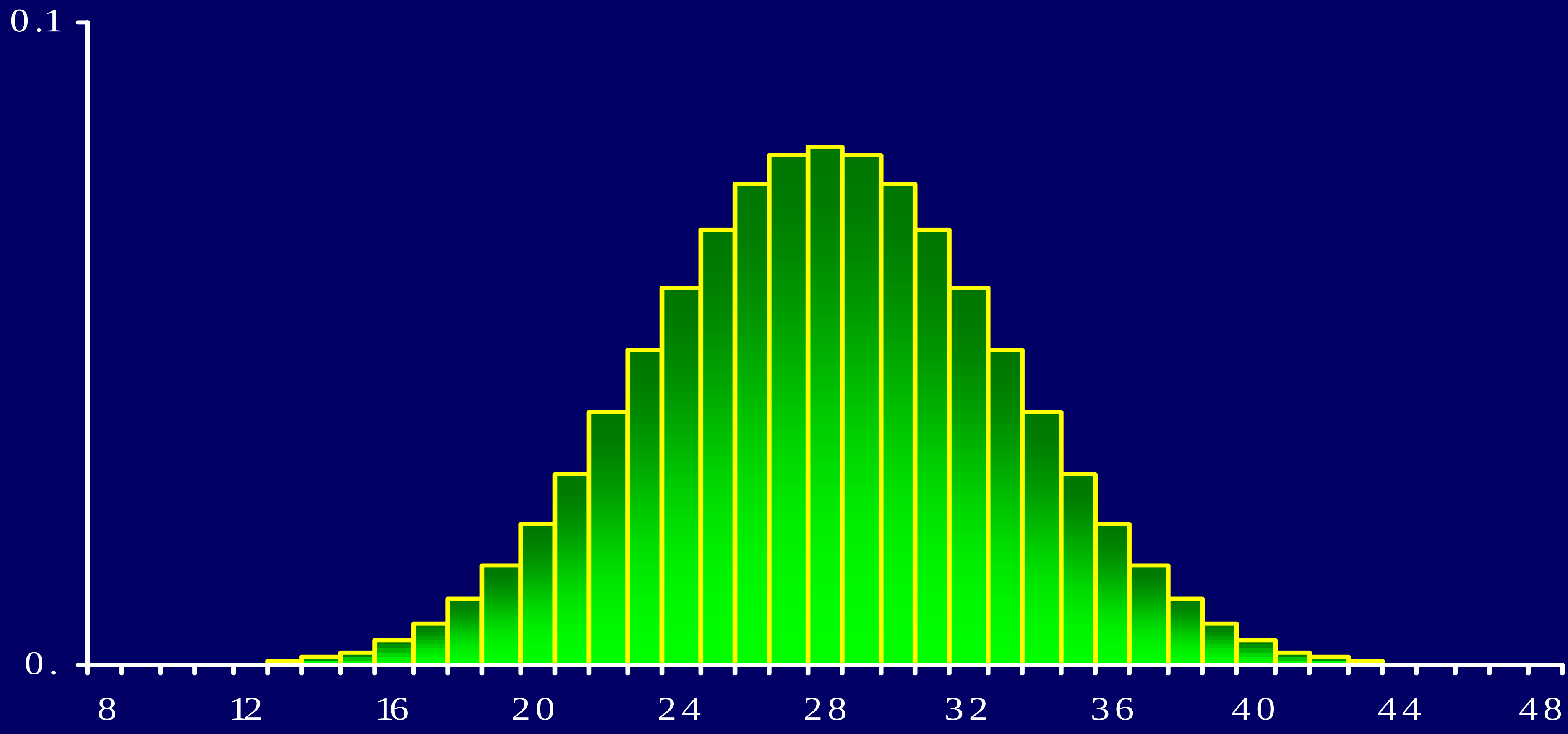
by independence of S_2 and S'_2

$$= \sum_{i=1}^{k-1} P(S_2=i)\, P(S'_2=k-i).$$

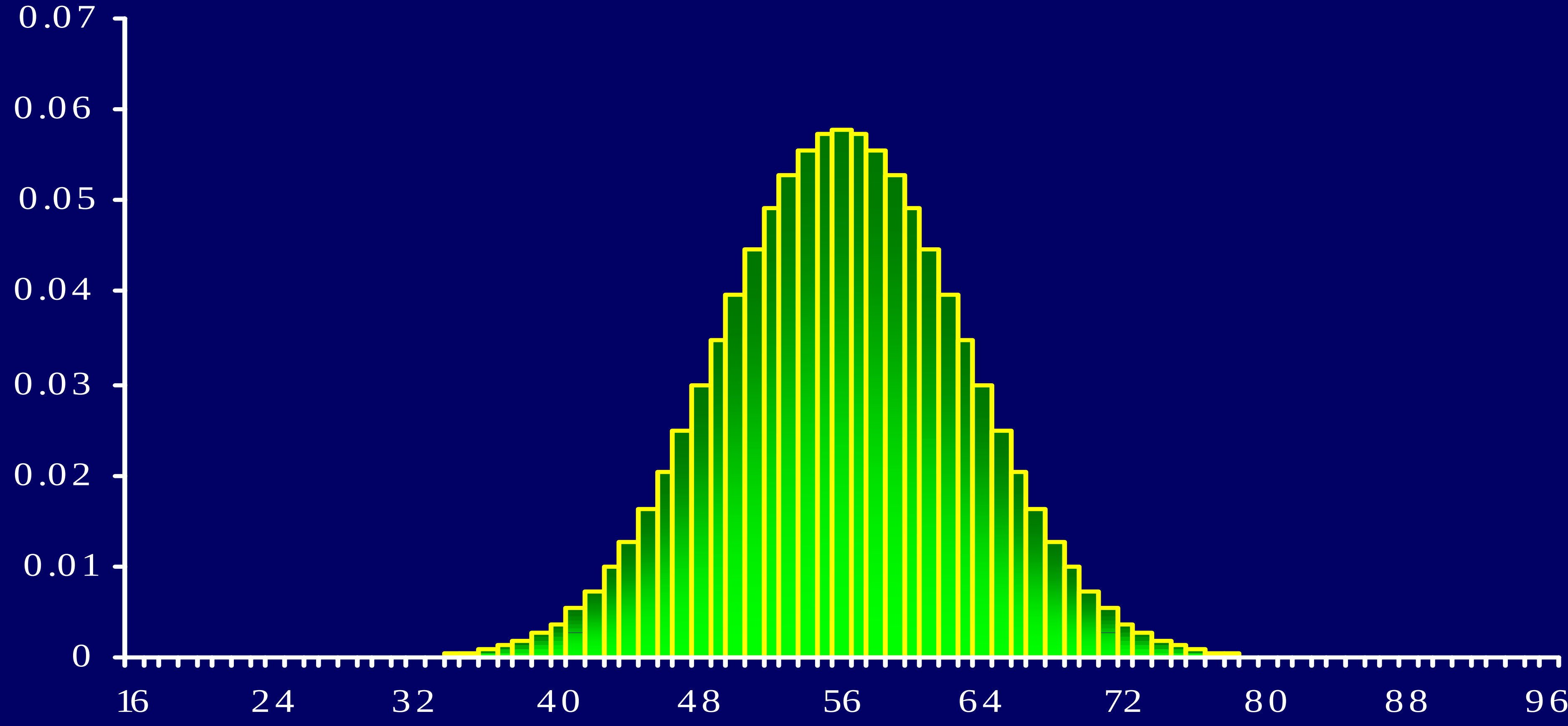
Distribution of S_4



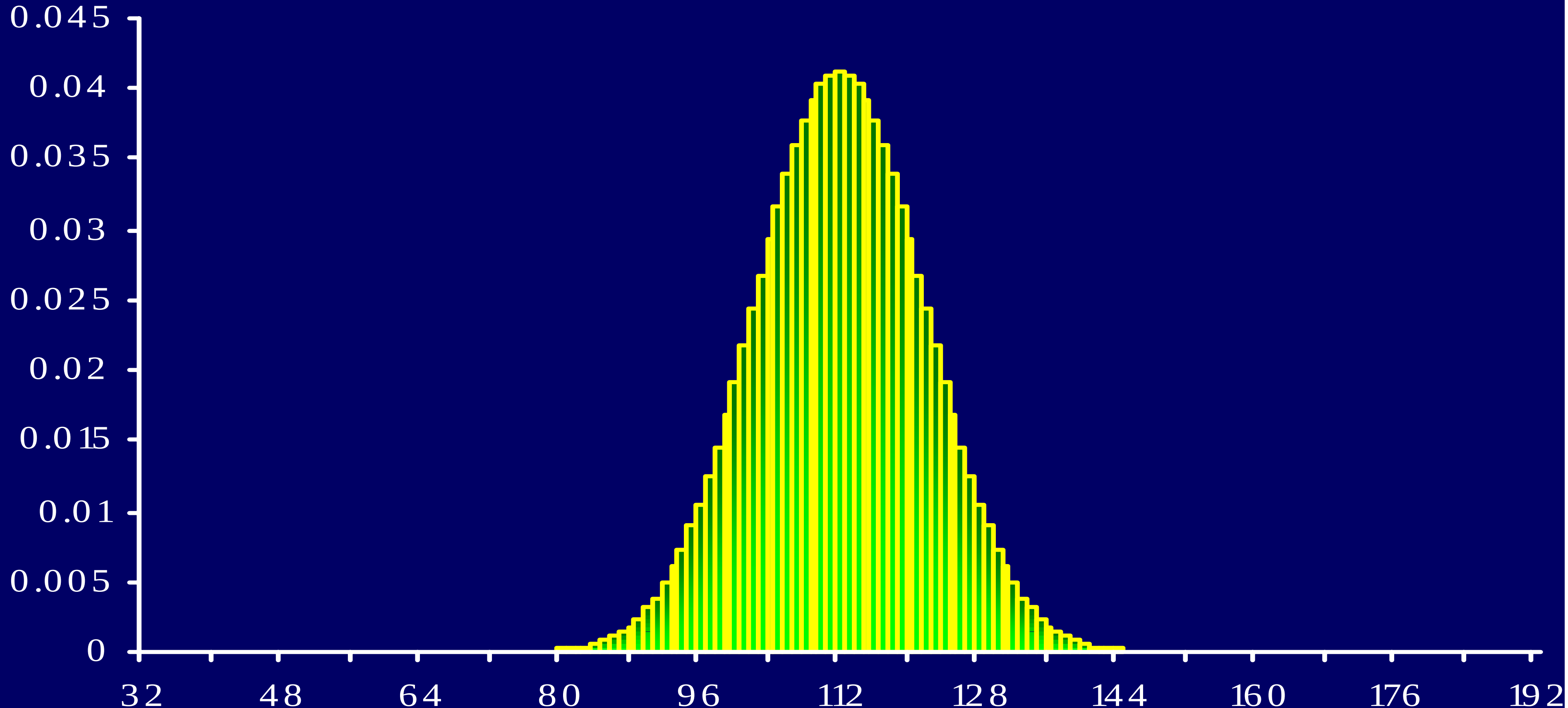
Distribution of S_8



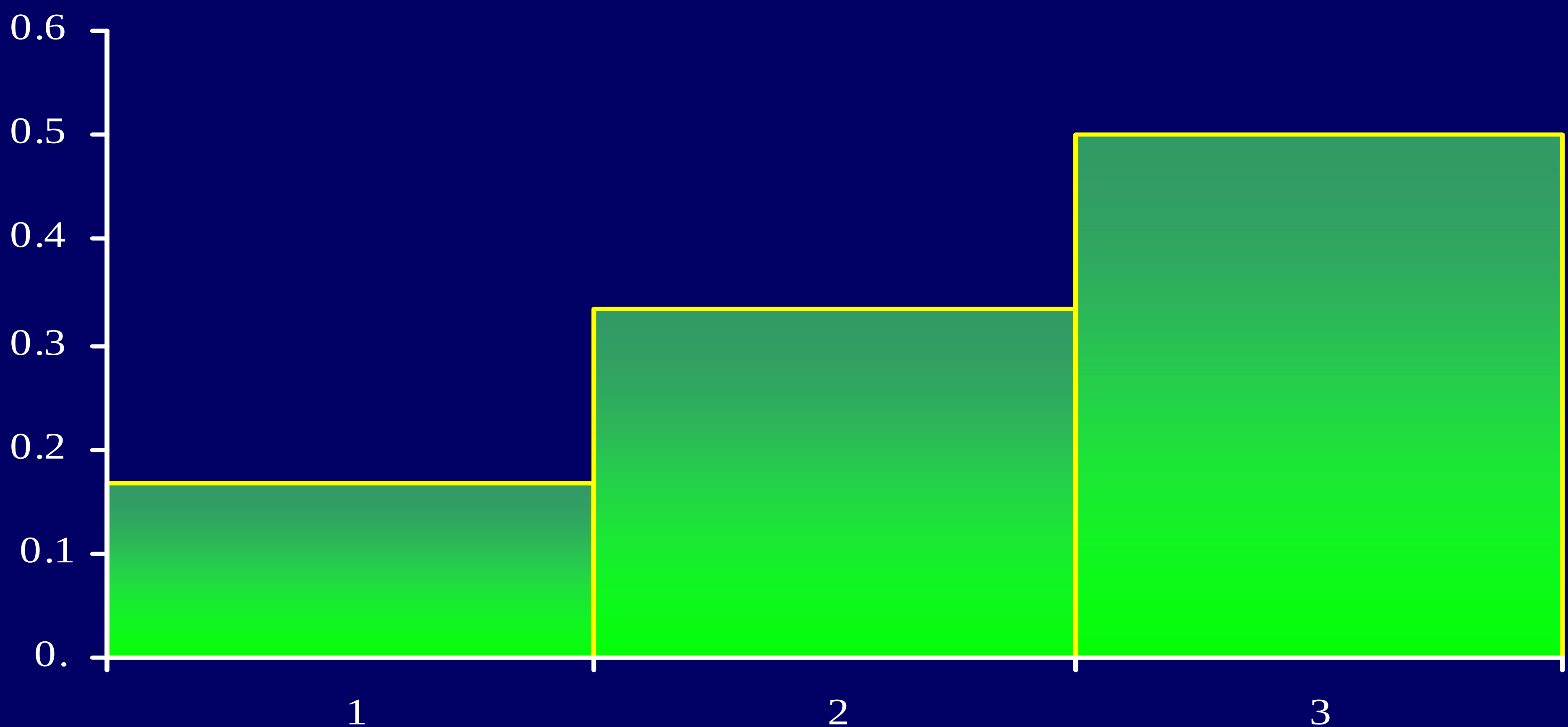
Distribution of S_{16}



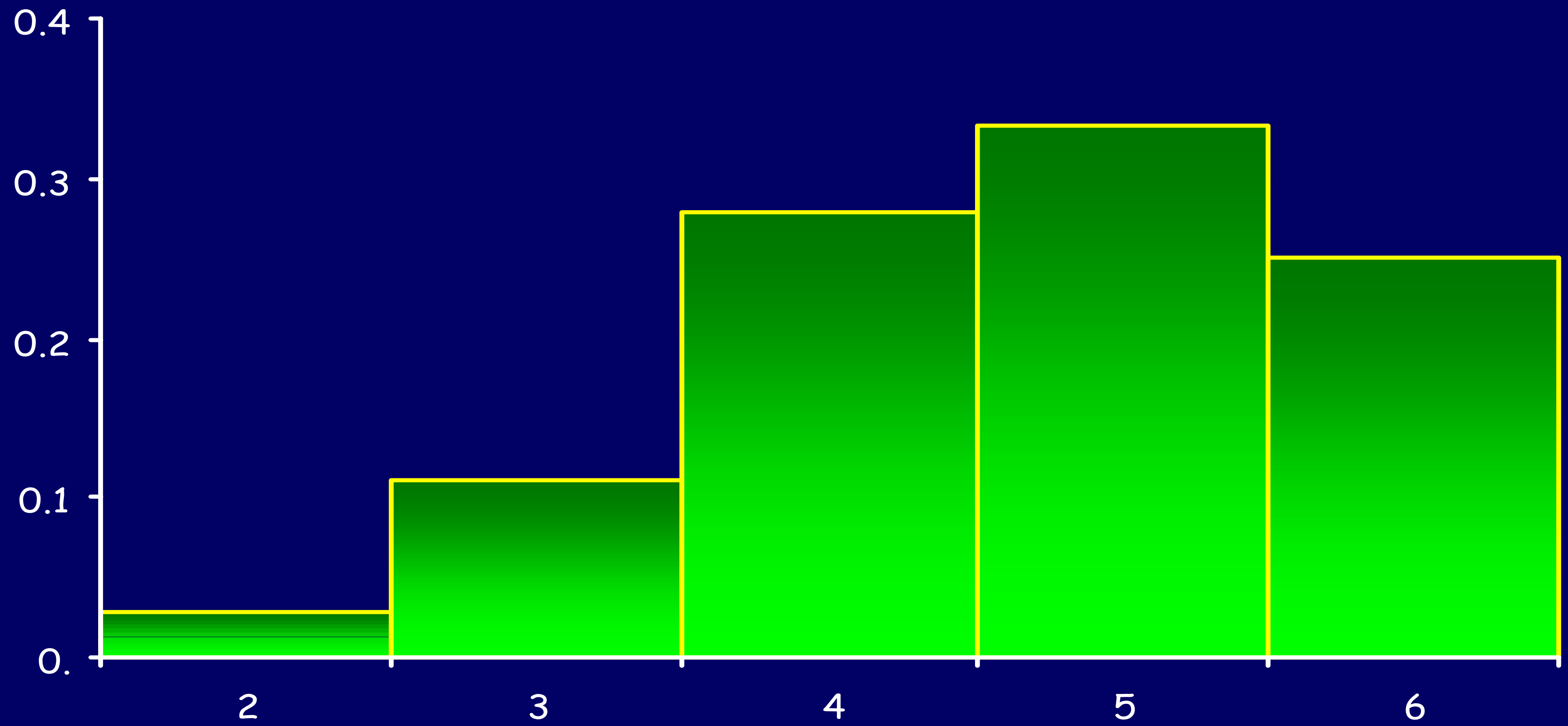
Distribution of S_{32}



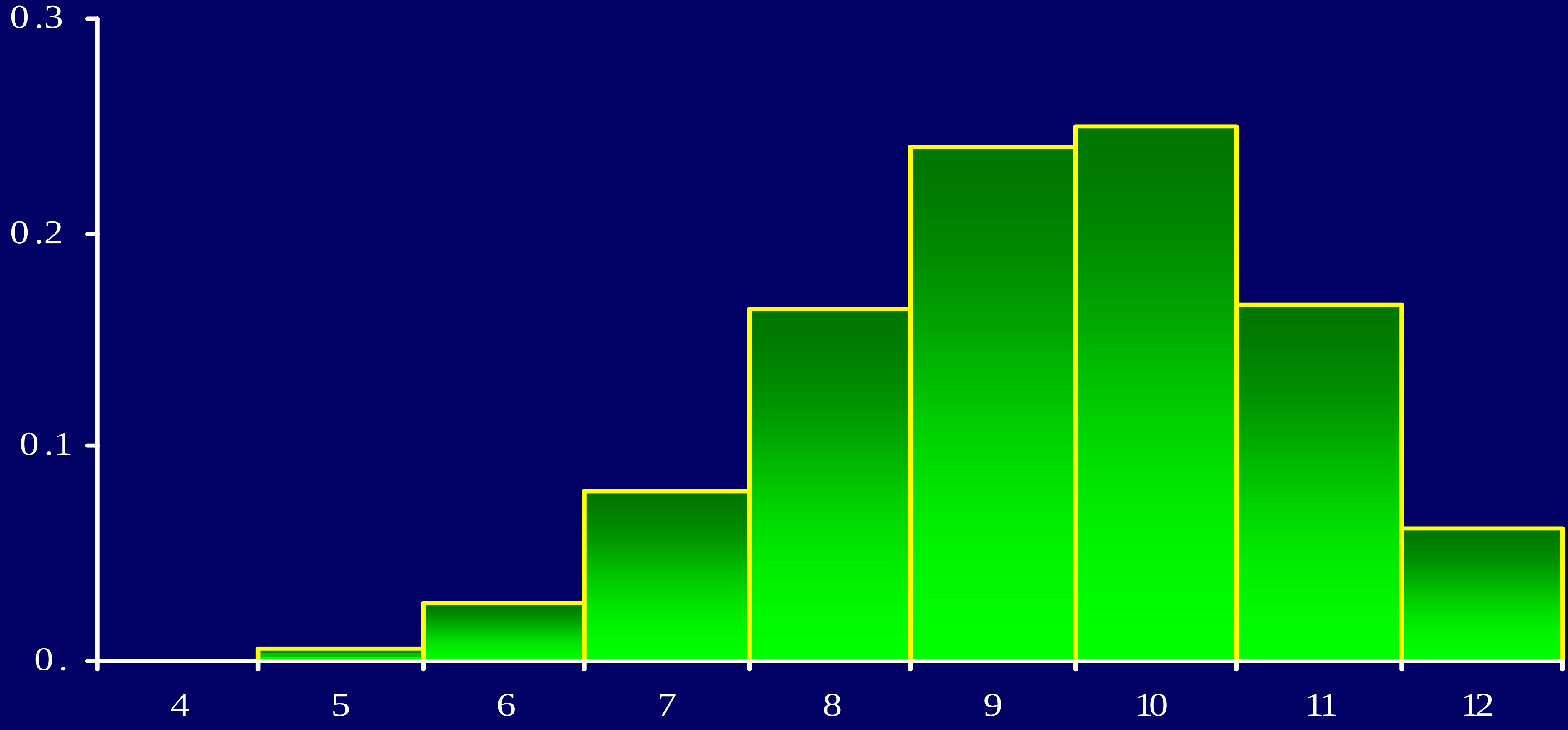
Distribution of X_1



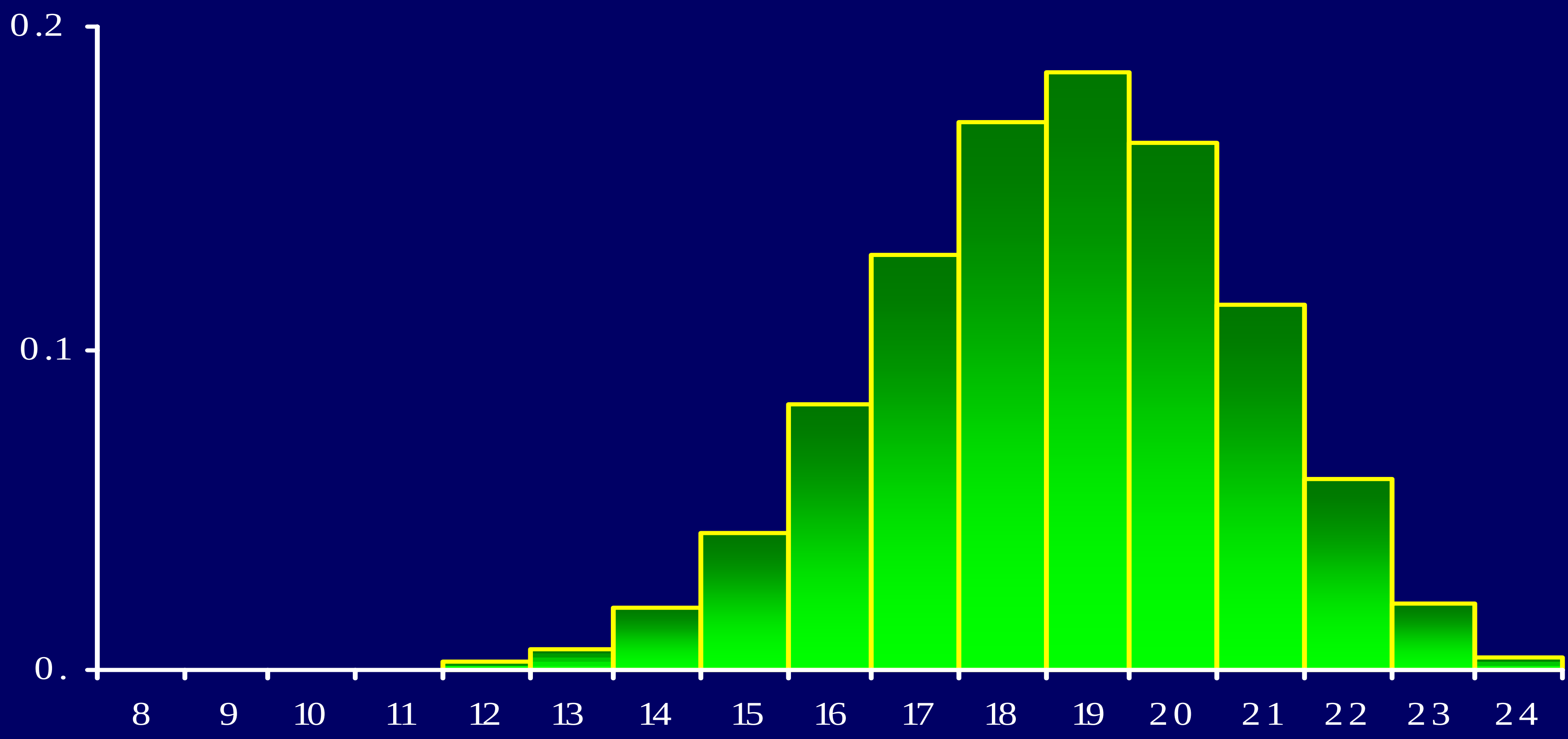
Distribution of S_2



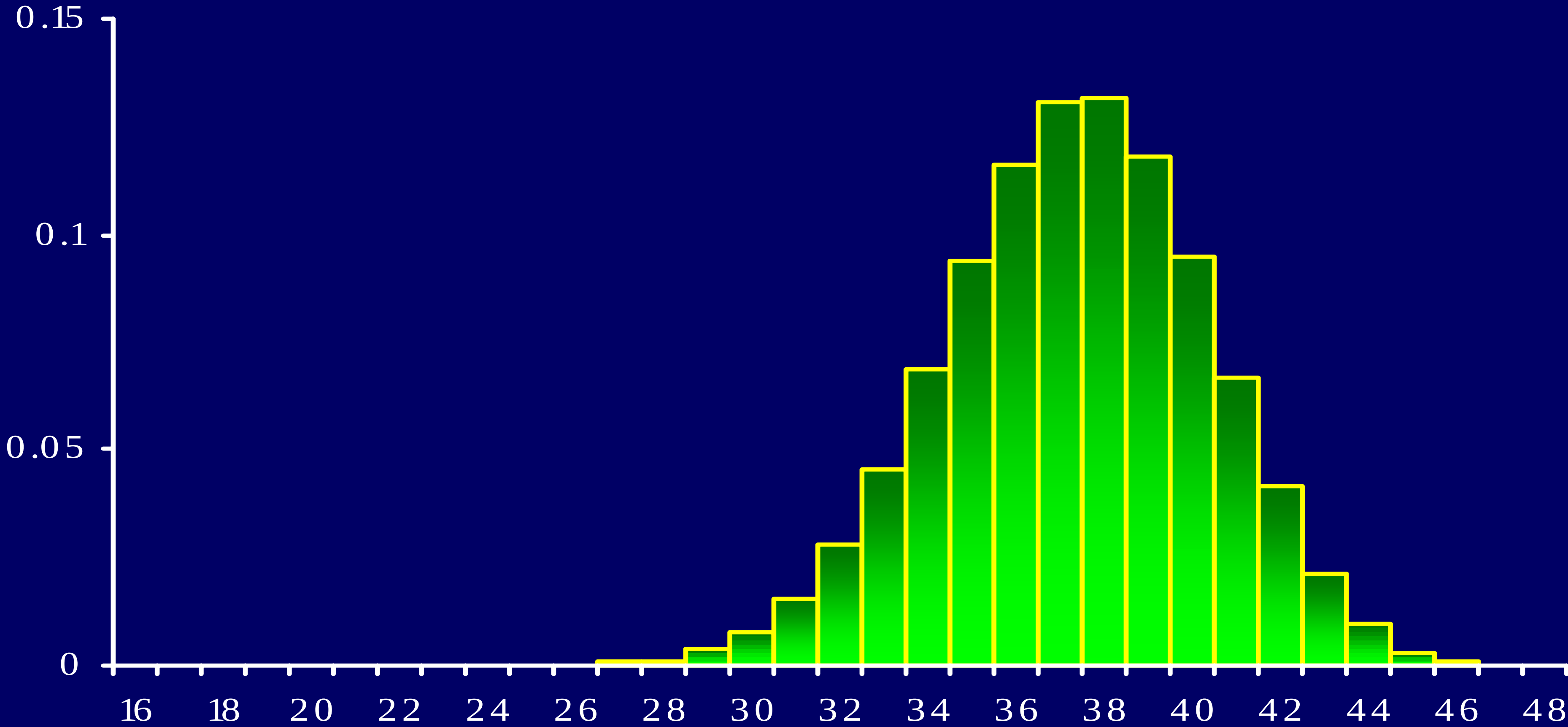
Distribution of S_4



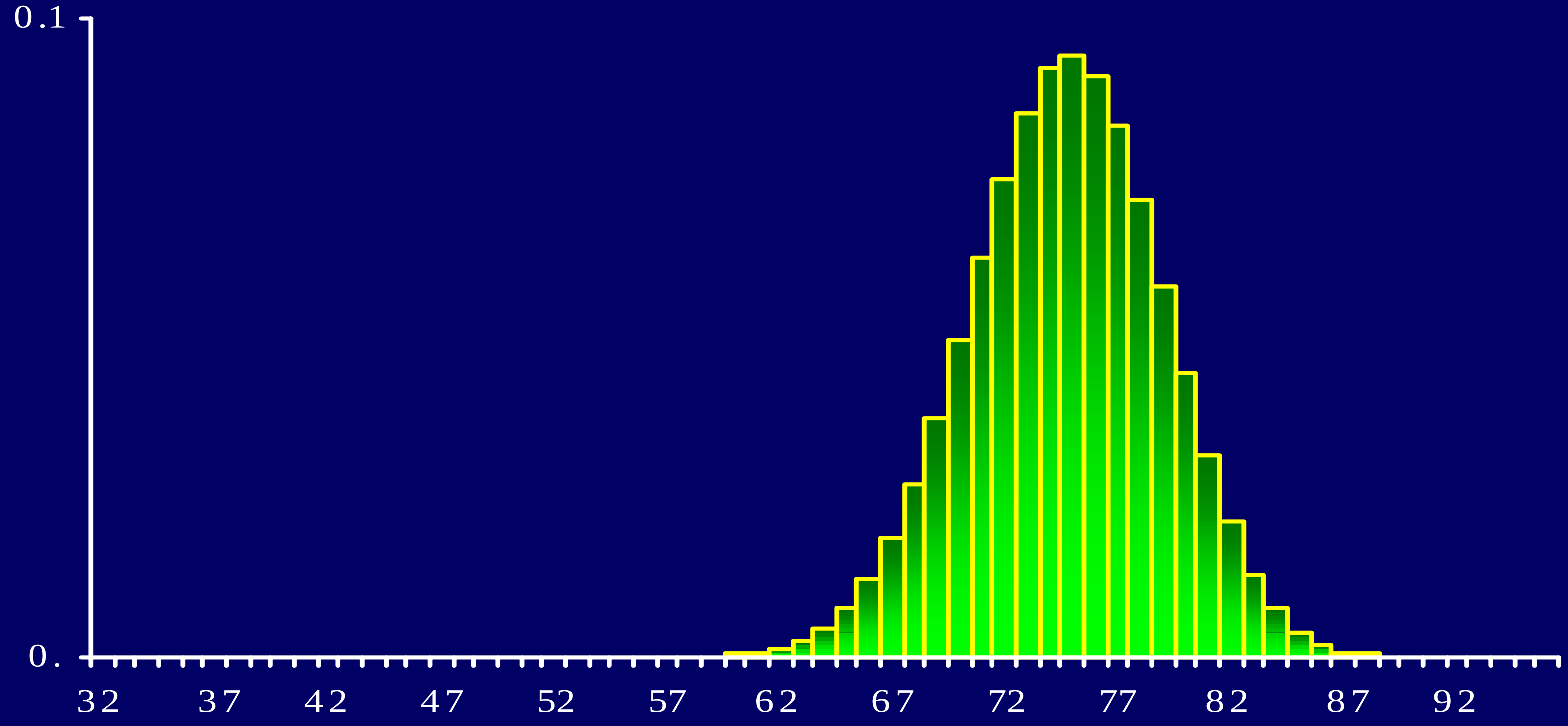
Distribution of S_8



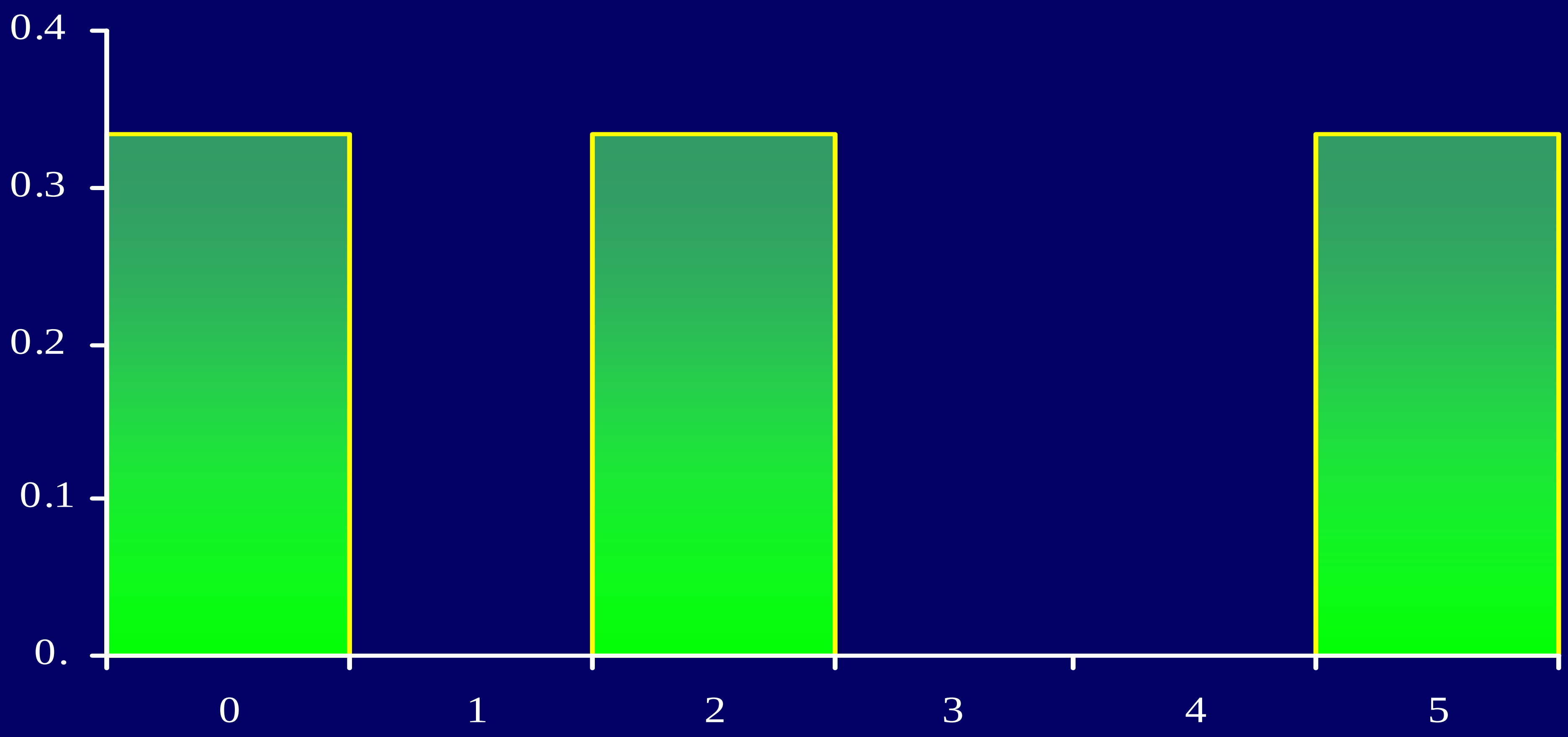
Distribution of S_{16}



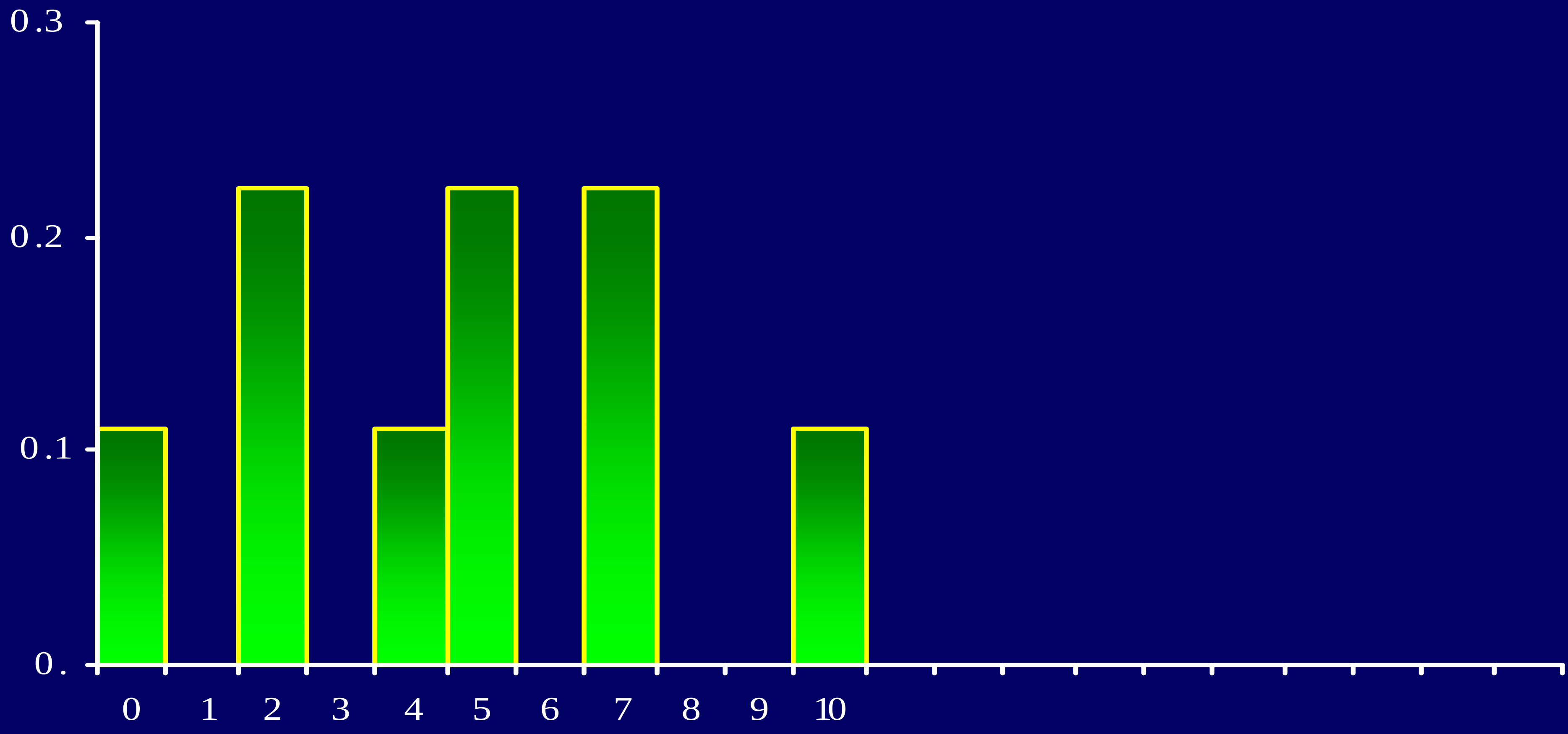
Distribution of S_{32}



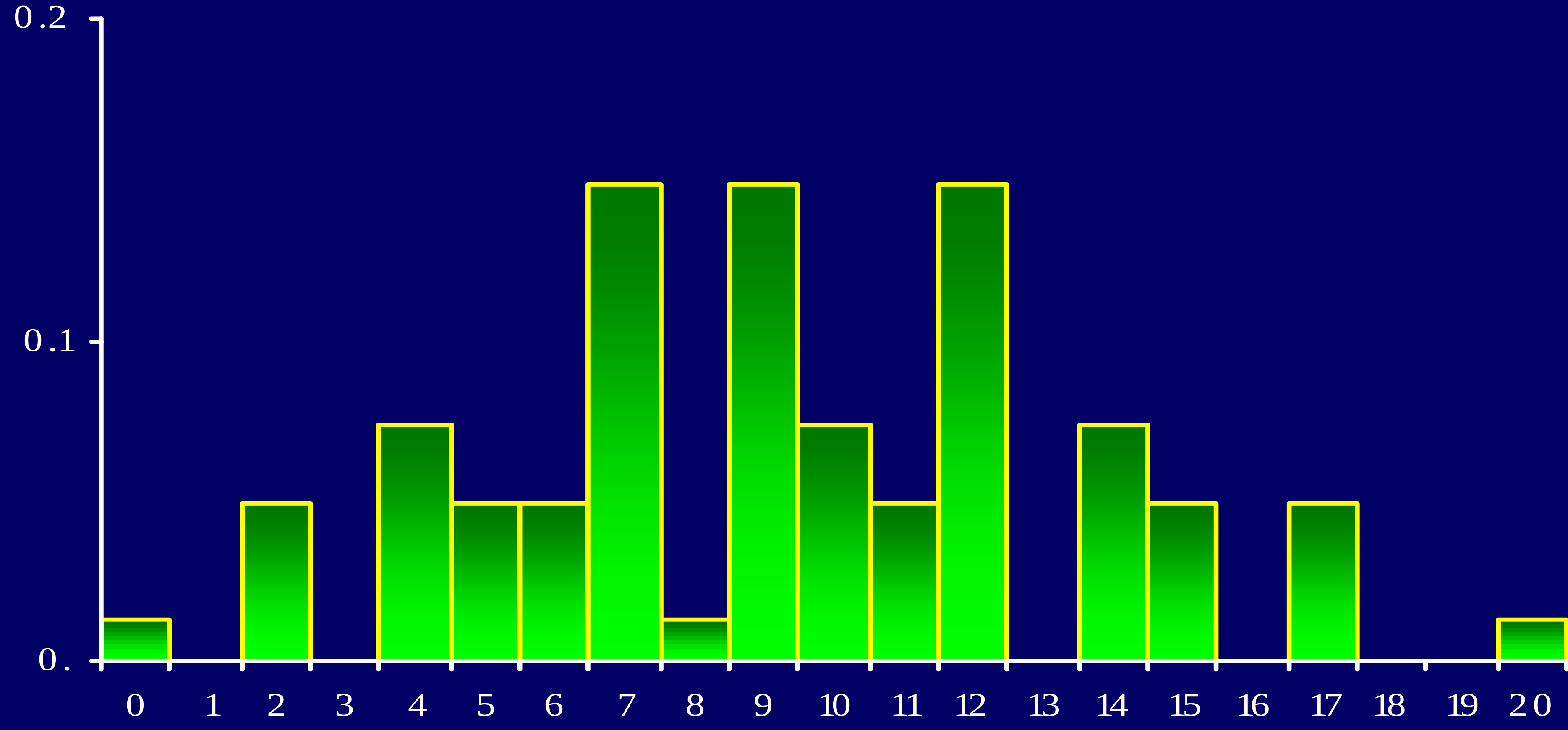
Distribution of X_1



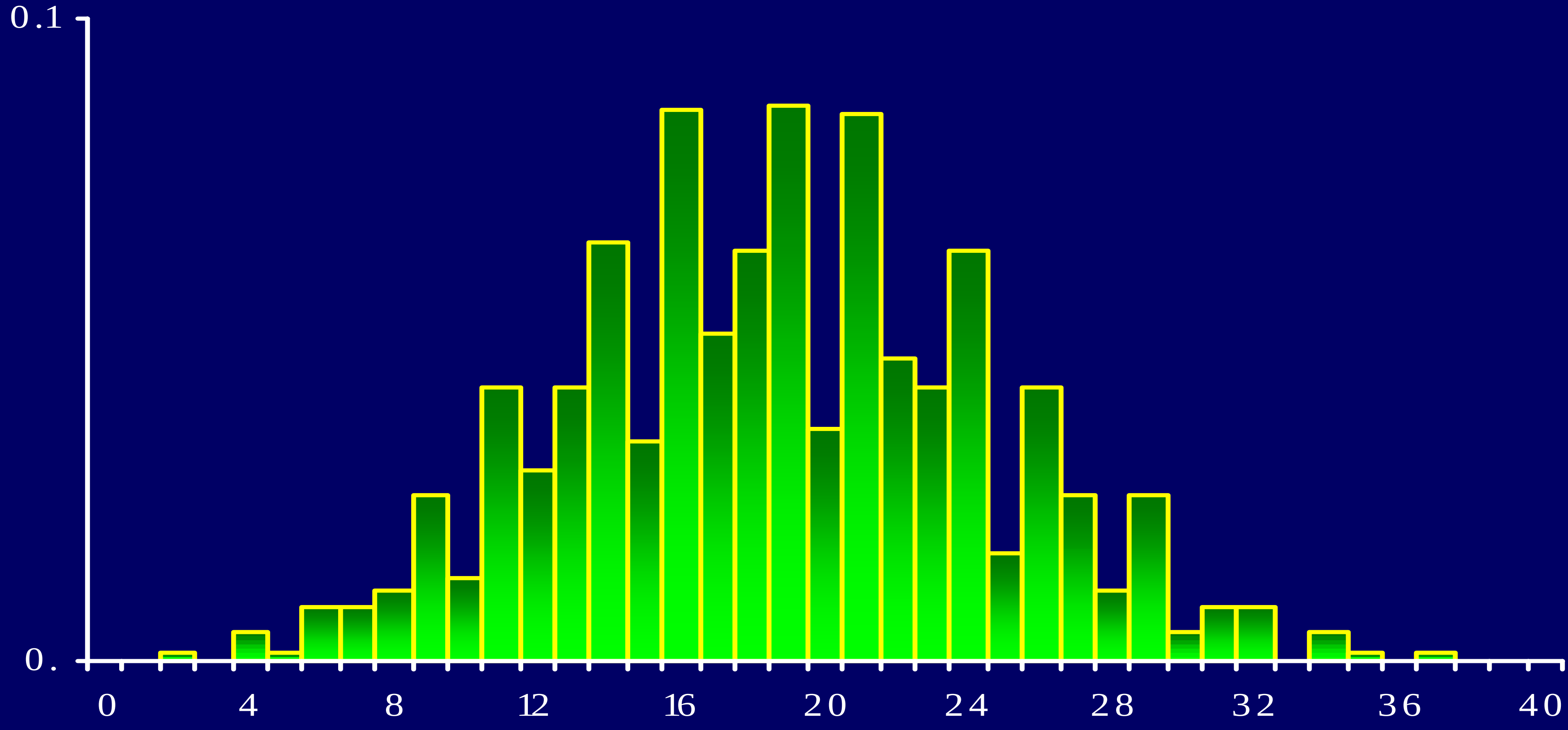
Distribution of S_2



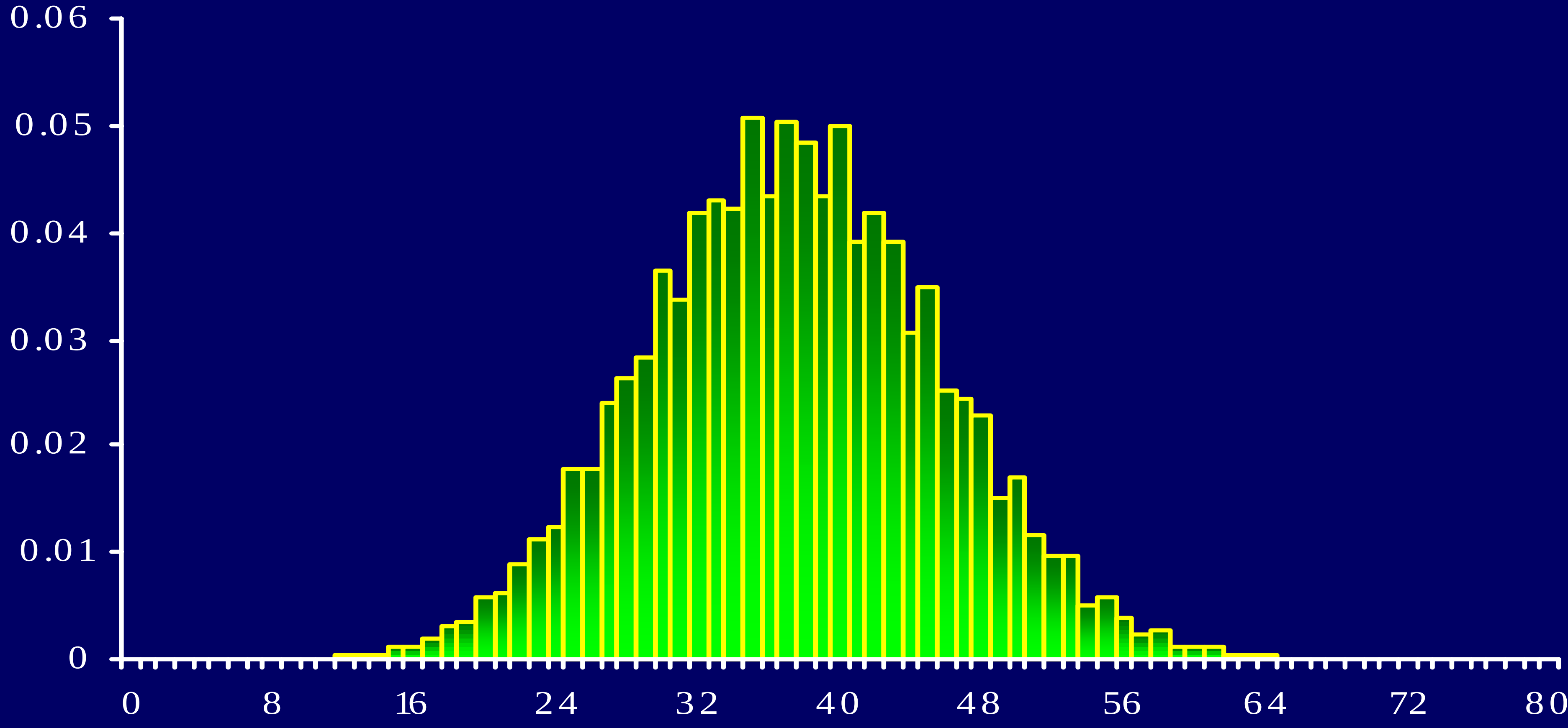
Distribution of S_4



Distribution of S_8



Distribution of S_{16}



Distribution of S_{32}